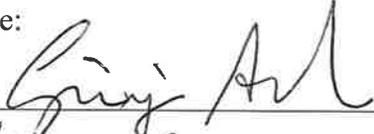


MORPHOLOGICAL CLASSIFICATION OF GLIA: A NEUROINFORMATICS
APPROACH

by

Masood Akram
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Bioengineering

Committee:











Dr. Giorgio Ascoli, Dissertation Director

Dr. Qi Wei, Committee Member

Dr. Caroline Hoemann, Committee Member

Dr. Patrick McKnight, Committee Member

Dr. Shani Ross, Department Associate Chair

Date: April 25, 2022

Spring Semester 2022
George Mason University
Fairfax, VA

Morphological Classification of Glia: a Neuroinformatics Approach

A Dissertation submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy at George Mason University

by

Masood A. Akram
Doctor of Pharmacy
University of Karachi, 2012

Director: Giorgio A. Ascoli, Professor
Department of Bioengineering

Spring Semester 2022
George Mason University
Fairfax, VA



This work is licensed under a
Creative Commons Attribution-Non-commercial (CC-BY-NC) v3.0
un-ported license.

DEDICATION

I dedicate my PhD dissertation to my parents M. Akram Arif and Mansoor Sharif, my sisters Madeeha Akram and Ayesha Akram, and my brother Mansoor A. Akram.

ACKNOWLEDGEMENTS

This project would not have been possible without the contributions of my colleagues, and mentor. To begin with, I would like to thank my mentor and advisor, Dr. Ascoli, for giving me this chance, guiding me on my long journey, teaching me new skills and polishing my overall abilities, believing in me, and always being so understanding and accommodating when it was difficult. Thanks to all my committee members for their constructive criticisms and unique contributions. My gratitude also goes out to all the lab members who made this journey so special.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	ix
List of Equations	xi
List of Abbreviations	xii
Abstract	xiii
Chapter 1: Introduction	1
Chapter 2: An open repository for single-cell reconstructions of the brain forest	10
Abstract	10
Introduction	11
Results	12
Database Content	12
Web Functionality	13
Usage Statistics	21
Representative Applications	23
Discussion	24
Methods	28
Author's contributions	33
Competing interests	33
Acknowledgements	33
Chapter 3: Efficient metadata mining of web-accessible neural morphologies	35
Abstract	35
Introduction	36
Materials and Methods	39
Results	44
Discussion	57
Acknowledgments	60
Funding	60
Chapter 4: Machine learning classification reveals robust morphometric biomarker of glia and neuronal arbors	61

Abstract.....	61
Introduction.....	62
Materials and Methods.....	65
Results.....	73
Discussion.....	93
Appendix 1: Genetic single neuron anatomy reveals fine granularity of cortical axo-axonic cells	100
Summary.....	100
Graphical Abstract.....	101
Introduction.....	102
Results.....	105
Establishing a gSNA Platform.....	105
AACs Tend to Localize at the Borders between Cortical Layers.....	112
AACs Elaborate Laminar-Restricted Apical and Basal Dendrites That Protrude Dendritic Spines.....	113
AACs Elaborate Laminar-Stratified Axon Arbors, Some with Translaminar Arbors.....	118
AACs Consist of Multiple Subtypes Distinguished by Dendrite-Axon Distributions That Reflect Input-Output Connectivity Patterns.....	123
AAC Subtypes Can Be Revealed by Axon Terminal Geometry That Correlates with That of Postsynaptic AIS.....	131
Discussion.....	134
STAR★Methods.....	139
Key Resources Table.....	139
Contact for Reagent and Resource Sharing.....	141
Experimental Model and Subject Details.....	141
Method Details.....	142
Quantification and Statistical Analysis.....	146
Data and Software Availability.....	154
Acknowledgments.....	154
Supplemental information.....	155
Appendix 2: An open-source framework for neuroscience metadata management applied to digital reconstructions of neuronal morphology.....	167
Abstract.....	167
Introduction.....	168

Methods	172
Organization of NeuroMorpho.Org metadata	172
Design and implementation of the metadata portal.....	176
Results.....	179
Metadata complexity, time saving, and error reduction.....	180
Usage protocol	186
Conclusion	194
Availability of data and materials	195
Acknowledgements.....	196
Appendix 3: Large scale similarity search across digital reconstructions of neural morphology	197
Abstract.....	197
Introduction.....	198
Materials and Methods.....	201
Results.....	206
Discussion.....	214
Acknowledgements.....	217
Appendix 4: scale correction	218
Glia.....	218
Neurons.....	220
Conclusion	224
References.....	228
Biography.....	280

LIST OF TABLES

Table 3. 1. Summary reporting use case scenarios	48
Table A1. 1. Key resource table.....	139
Table A1S1. 1. Complete cell list and Neurolucida dataset on morphometry.....	165
Table A2. 1. Number of distinct scientific concepts.....	182
Table A3. 1. Duplicate detection false positive rate (FPR)	213

LIST OF FIGURES

Figure 2. 1. Digital reconstructions of cellular morphologies in NeuroMorpho.Org.	16
Figure 2. 2. Browse functionality.	17
Figure 2. 3. NeuroMorpho.Org’s API gives developers and data miners access to database content and information allowing integration with other neuroinformatics tools.	20
Figure 2. 4. Infographics summary of data download activity from the NeuroMorpho.Org portal.	20
Figure 2. 5. Flowchart of the operational procedures of NeuroMorpho.Org.....	27
Figure 2. 6. Literature database and search.....	30
Figure 3. 1. Growth and complexity of data in NeuroMorpho.Org.....	38
Figure 3. 2. Data, metadata, and morphometry content of NeuroMorpho.Org.	41
Figure 3. 3. Summary reporting graphical user interface.	45
Figure 3. 4. Metadata and morphometry summaries.	51
Figure 3. 5. Group distinctiveness and feature distinguishing power of NeuroMorpho.Org glia and monkey principal neurons.	55
Figure 4. 1. Representative diversity of morphological reconstructions of glia and neurons from NeuroMorpho.Org with labels indicating animal species, anatomical region, and cell type.	74
Figure 4. 2. Balanced distribution.....	76
Figure 4. 3. Schematic of selected morphometric features.	78
Figure 4. 4. Orthogonalization of morphometric features.	80
Figure 4. 5. PCA biplot.....	82
Figure 4. 6. Classification performance for Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF), including the area under the curve (AUC) of the Receiver Operating Characteristic plot.	84
Figure 4. 7. Silhouette profiles of length, height, contraction, number of branches, and average branch Euclidean length (ABEL) of glia and neurons	86
Figure 4. 8. Classification performance of average branch Euclidean length (ABEL).	88
Figure 4. 9. Relationship between the average branch Euclidean length (ABEL) of terminal branches and internal (bifurcating) branches for glia (blue) and neurons (green).	90
Figure 4. 10. Classification of glia and neurons across anatomical regions.	92
Figure A1. 1. Schematic of the gSNA Platform Applied to the Mouse Brain.....	106

Figure A1. 2. Areal and Laminar Distribution of AACs Revealed from Whole-Brain fMOST Dataset.	108
Figure A1. 3. Single AAC Reconstructions in Cerebral Cortex.	110
Figure A1. 4. Characteristics of L2 AAC Dendrites.	115
Figure A1. 5. Morphology and Distribution Patterns of AAC Axons in the Neocortex.	121
Figure A1. 6. Hierarchical Clustering of AACs Based on Cortical Laminar Density Distribution of Axons and Dendrites.	126
Figure A1. 7. AAC Subtypes Revealed by Axon Terminal Characteristics That Correlate with AIS.	133
Figure A1S1. 1. Sparse labeling of AACs in mPFC, MC and SSC.	156
Figure A1S1. 2. AAC reconstructions (orientation adjusted) and typical AAC individuals.	157
Figure A1S1. 3. Characteristics of L2 AAC dendrites.	159
Figure A1S1. 4. Cylindrical length density distribution pattern of L2 AAC axons.	160
Figure A1S1. 5. Representative individual and clustered AACs in the neocortex.	162
Figure A1S1. 6. Schematic of normalizing axonal and dendritic length-density distribution to a standard cortex template.	164
Figure A1S1. 7. Clique analysis of AAC types.	164
Figure A2. 1. Metadata organization in NeuroMorpho.Org.	176
Figure A2. 2. Overview of the system's architecture.	177
Figure A2. 3. Labor-saving per version release yielded by the metadata portal.	186
Figure A2. 4. View of the portal's main page.	188
Figure A2. 5. Newly created dataset in the metadata portal.	190
Figure A2. 6. Metadata form to annotate the details of the reconstruction within each experimental group.	192
Figure A2. 7. Final dataset with reconstructions and all experimental groups added in the metadata portal.	193
Figure A3. 1. Similarity search results compared to random controls for six representative cells from NeuroMorpho.Org.	208
Figure A3. 2. Similarity search with the graphical user interface.	210
Figure A3. 3. Ratio of variance of each eigenvalue for the three different descriptors used in similarity search.	211

LIST OF EQUATIONS

Equation 4. 1	72
---------------------	----

LIST OF ABBREVIATIONS

API.....	Application programming interface
CMS.....	Content management system
HTML.....	Hyper text markup language
NIF.....	Neuroscience information framework
OWL.....	Web ontology language
PDF.....	Portable document format
REST.....	Representational state transfer
RW.....	Read and write
SQL.....	Structured query language
UI.....	User interface
URL.....	Uniform resource locator
XML.....	eXtensible markup language

ABSTRACT

MORPHOLOGICAL CLASSIFICATION OF GLIA: A NEUROINFORMATICS APPROACH

Masood A. Akram, Ph.D.

George Mason University, 2022

Dissertation Director: Dr. Giorgio A. Ascoli

Neurons and glia are the two main types of cells in the nervous system. Neurons communicate by transmitting signals giving different species the ability to perform various complex functions. As opposed to neurons, glial cells support the nervous system and are responsible for maintaining normal homeostasis. Glial cells are as abundant as neurons in the nervous systems of most animals, including humans. Just like neurons, glial cells are also characterized by complex branching morphologies. Scientists have long been interested in classifying neurons; however, little attention has been given to classifying glia.

In recent years, advances in neuroscience research have led to an increased interest in neuroinformatics, data sharing, and online data repositories, as well as a related need for data organization. NeuroMorpho.Org is the world's largest public repository of digitally reconstructed neural morphology with more than 172,000 traced cells. This

online neuralbank provides open access to neural reconstructions so that published data can be downloaded for statistical and mathematical modeling, leading to new scientific discoveries. Each tracing entry in NeuroMorpho.Org is accompanied by a battery of morphometric features such as length, volume, angles, diameter, etc., as well as detailed metadata annotations describing the animal subject, anatomy, and experimental preparation.

As a member of the NeuroMorpho.Org team, I have mastered different tasks including database maintenance, and helped NeuroMorpho.Org grow from 62,000 to more than 172,000 digital neural tracings, of which glia now constitute more than 10% of the total content. However, content expansion and continuous growth increased the complexity of data and metadata, requiring effective resources for information access. Therefore, we introduced the online tool *Summary reporting* to generate structured reports of morphometry organized into homogenous metadata groups for arbitrary subsets of data selected by the user.

As the first application of this new functionality, we focused on advancing models to classify neurons and glia. We applied supervised learning algorithms including Support Vector Machine, Random Forest, and K-Nearest Neighbors to distinguish neurons and glia based on their extracted morphometric features. Across a diverse set of metadata, including species, brain regions, and histological processing, the classifiers were able to discern neurons and glia with high precision. Our results indicate that arbor morphology is an effective and robust way to categorize brain cells. In particular, our

work identifies an individual morphometric measure, Average Branch Euclidean Length (ABEL), which can be robustly used to distinguish neurons from glia across different vertebrate and invertebrate animal models, a variety of experimental conditions, and anatomical regions, except for the cerebellum. In addition, we found that calculating the ABEL from as few as five branches can provide above 95% accuracy in classifying neurons and glia without requiring the entire cell arbor to be reconstructed.

CHAPTER 1: INTRODUCTION

Cellular classification has been a long-studied topic since the inception of neuroscience, because it provides the fundamental parts list underlying brain structure and function: in other words, it strives to identify the building blocks defining the key neural correlates of learning, memory and cognition (Armañanzas & Ascoli, 2015a). Morphology, electrophysiology, and biochemistry are the most common methods used to study cells in the nervous system (The Petilla Interneuron Nomenclature Group (PING), 2008). Over the last decade, these three methods have witnessed tremendous technological breakthroughs (Litvina et al., 2019a). In particular, morphological studies have benefited from substantial advancement in microscopy, tracing software, computational modeling, and machine learning (Peng et al., 2017a). However, neuroscientists have traditionally focused their classification techniques on neurons (Bota & Swanson, 2007a); very few studies have been undertaken to develop an integrated classification of glia, despite their similar abundance to neurons (von Bartheld et al., 2016). Therefore, understanding the morphological features that distinguish glia from neurons and the best method for categorizing the significant types of glia is an essential but unfulfilled step in developing a complete cell census of any nervous system.

Neurons and glia are the main cellular components of nervous systems. Neurons transmit and process information via action potentials, whereas glial cells maintain

homeostasis within the nervous system. The three major types of neurons are principal cells, interneurons, and sensory cells. A principal cell is a long-range projecting neuron (Shepherd et al., 2019) that transmits electrical signals to non-adjacent anatomical regions, allowing brain-wide information integration. Local interneurons are the neurons whose actions, excitation or inhibition, are limited to the local circuit (Shepherd et al., 2019). Sensory neurons are responsible for receiving a range of chemical and physical stimuli, including temperature, pain, and touch, from the environment and various organs, and transmitting them to the brain (Donnelly et al., 2020). Like neurons, glial cells are also divided into multiple types. Among the main types of glia are microglia, astrocytes, and oligodendrocytes. Microglia are protective cells in the brain and cause inflammation in response to injury or disease. In addition, they play an essential role in suppressing excess synaptic connections during development (Aguzzi et al., 2013a). Astrocytes regulate neurotransmitter levels in the brain, support synaptic modification (Ullian et al., 2001), and communicate with neurons via calcium signals (J. Yang et al., 2016). They are the most abundant glial cells in the central nervous system (CNS) and play an essential role in the formation of the blood brain barrier, regulation of blood flow, synaptic, circuit, and behavioral functions (Zhou et al., 2019). The oligodendrocytes provide myelin insulation to the long-range projection axons (Nave, 2010) in the central nervous system to minimize the electrical leaking of impulses. Additionally, oligodendrocytes aid in transferring energy from blood cells to axons.

The usage of confocal microscopes in conjunction with computers enabled the routine reconstruction and archiving of complex branching morphologies (such as

dendrites, axons, and glial processes) as digital files containing three-dimensional coordinates as well as their tree structures (Senft, 2011a). Recent advances in microscopy and digital tracing have enabled a growing number of cellular reconstructions of neurons and glia to be uploaded to NeuroMorpho.Org, which is currently the largest repository of cellular neuronal and glial reconstructions in neuroscience. In NeuroMorpho.Org, each reconstruction is described in detail with metadata, providing a qualitative description of the cell (Parekh et al., 2015a). This information consists of 35 distinct fields pertaining to the animal subject, anatomy, completeness, experimental procedure, and provenance. Most of this information is directly provided by the research groups who deposit their digital tracings, and then supplemented by in-house curators through the mining of the reference peer-reviewed articles. Besides the reconstruction files and metadata, each cell is also assigned with 21 standard morphometric parameters. L-Measure extracts these parameters, including the average branch diameter, number of branches, total arbor length, volume, bifurcation angles, topological asymmetry, fractal dimension, and taper rate, among others (Scorcioni et al., 2008b). An additional characterization of arbor geometry was recently added to NeuroMorpho.Org, called Persistence Diagram Vectors (PDVs). In PDVs, branching distributions of neuronal and glial trees are represented using the mathematical formalism of algebraic homology, which encompasses the information contained in popular Sholl diagrams – namely, the number of branch points by Euclidean distance from the soma (Y. Li, Wang, et al., 2017a).

Founded in 2006 with an initial collection of fewer than 1000 neurons (Ascoli et al., 2007a), NeuroMorpho.Org provides unhindered access to all digital tracings of neuronal morphology freely shared upon request by researchers. Glial cells were added to this online repository in version 7.1, released in 2017. With version 8.1.137 (the latest at the time of this writing: released March 21st, 2022), glial cells constitute about 11% of more than 170,000 cellular reconstructions. Hundreds of laboratories worldwide continuously contribute neurons and glia tracings to NeuroMorpho.Org, which are processed, morphometrically quantified, and annotated with comprehensive metadata. During my doctoral studies, I had the opportunity to learn about, and contribute to, different aspects of the maintenance and continuous development of NeuroMorpho.Org. Chapter 2 describes the four main stages of this process: literature mining, data processing, metadata curation, and public release. Literature mining involves searching and shortlisting published articles based on an automated tool (Maraver et al., 2019b), which are then thoroughly examined. If digital tracings are found in the published articles, authors are contacted to share their data for public posting. The metadata is typically provided by the author, but in some instances, the annotations are obtained from the published article by the curators. The digital tracings added to the database must meet specific criteria, so the shared files from authors are checked using different software tools, such as CVapp (Cannon et al., 1998c), Neuromantic (Myatt et al., 2012), neuTube (Feng et al., 2015), and several others (Parekh & Ascoli, 2013), to make sure they are at the required standard. Once the reconstructions have met the standards, they are added to

NeuroMorpho.Org in the pre-release version and authors are asked to review the online archive before public release (Akram et al., 2018b).

Due to the rapid expansion of the NeuroMorpho.Org data as more and more labs around the world share their experimental data, numerous functionalities have been continuously added to NeuroMorpho.Org since its inception to make data reuse efficient and user-friendly. Among these features are early morphological pipelines (Halavi et al., 2008a), the ontological search engine (Polavaram & Ascoli, 2017a), and more recent advances in extended file standards (Nanda et al., 2018b) and automated literature mining (Maraver et al., 2019b). In addition, the Application Programming Interface (API) can be utilized in order to download the data in bulk from the online repository. Despite this, a graphical user interface that is intuitive and interactive was missing, which is an undeniable obstacle for non-technical users. By introducing Summary Reporting (Akram et al., 2022a), we have provided researchers with novel data mining capabilities, which are described in chapter 3. Specifically, this new functionality provides the ability to generate and download comprehensive metadata and morphometric features for arbitrary subsets of content within the database via web-accessible dropdown menus. In addition to the valuable resources already available within the neuroinformatics' workbench, this new open-source resource will help end-users identify the most appropriate datasets for their research questions.

In chapter 4, Summary Reporting was utilized to download metadata and

morphometrics of a balanced dataset of glia and neurons for the purpose of applying three supervised learning classifiers: Random Forest (Breiman, 2001a, 2001a; Sarica et al., 2017b), K-Nearest Neighbor (Aha et al., 1991a), and Support Vector Machine (Cortes & Vapnik, 1995a). All three algorithms performed exceptionally in classifying glia and neurons. However, recognizing the morphological signatures distinguishing glia from neurons was an important yet unfulfilled step. We investigated different morphometric features, such as overall arbor height and length, maximum Euclidean distance, average contraction, total volume, number of branches, etc., to understand how they contributed to classification. Previous evidences suggested that the height, length, volume, and tortuosity are all useful parameters for classification (Matas et al., 2013). Furthermore, it is believed that neurons are bigger than glia (García-Marín et al., 2007a; Y. Lu et al., 2015a; Veldman et al., 2020a; Zisis et al., 2021a), and that some glia are more complex than other types of glia (Khakh & Deneen, 2019; Verkhratsky et al., 2019). Based on the analysis of the machine learning results, we defined a new morphometric parameter, the Average Branch Euclidean Length (ABEL), and calculated it for each cell using the L-Measure (Scorcioni et al., 2008b). Our results showed that ABEL gave us better classification accuracy than any other morphological measure previously tried. Traditionally, neuron and glia have to be traced before different morphometric features could be measured. Some cells have as many as 4000 branches, making this process tedious and inefficient. Our analysis demonstrated that not every branch is required to categorize a cell; in fact, if only five branches are used, we can achieve an accuracy of over 95%.

In addition to the main chapters of this dissertation, I also participated in different projects mentioned in the appendices including the connectivity-guided morphological analysis of axo-axonic cells (appendix 1). Morphology is an intuitive presentation of various types of neurons, reflecting their connectivity (Seung & Sümbül, 2014; Sümbül et al., 2014). Even though the shape of a single neuron is not sufficient to identify all presynaptic sources and postsynaptic targets, it can be analyzed morphologically in a way that reflects global and local connectivity patterns. In this study, we examined the morphology of dendrites and axons of axo-axonic cells (chandelier cells), which are critical to network connectivity and cell function. This morphological analysis found that the differences in input-output connectivity were significant factors in determining the subtypes of neurons (Wang et al., 2019).

As an active member of the NeuroMorpho.Org team, I was involved in using, testing, and providing feedback on the metadata portal's functionality. In the past, metadata curation was purely manual, but a user-friendly online metadata portal that is less error-prone and labor-intensive has recently been developed (Bijari, Akram, et al., 2020b). The goal of this metadata portal is to allow data contributors from all over the world to share their neural reconstruction data as well as the metadata annotations quickly and efficiently. As a result, we reduced the labor costs, had fewer backlogs of data, and were able to provide the scientific community with rapid access to digital reconstructions for statistical and computational modeling. (appendix 2).

Moreover, I also contributed to a novel tool called Similarity Search (Ljungquist et al., 2021), as detailed in appendix 3. By means of this functionality, users can shortlist hundreds of thousands of neurons and glia that have similar structure and similarity score based on the selection of one cell quickly and efficiently by using the Facebook AI Similarity Search (FAISS) software. Cells are shortlisted based on the L-Measure derived morphometric features, persistence diagram vectors, and a combination of the two, by analyzing the explained variance following Principal Components Analysis (PCA) application.

Appendix 4 is comprised of the supplementary material of some archives analyzed in chapter 4. Most reconstructions in NeuroMorpho.Org have coordinates reported in microns, but there were some archives in which the coordinates were expressed in pixels. Therefore, we manually calculated the height of at least one cell in each archive and compared the resulting value to the height reported by NeuroMorpho.Org. If the values did not match, we revised the size-related morphometric features (width, height, total length, total surface, total volume, maximum Euclidean distance, and maximum path distance) of those archives.

In summary, in my tenure as a PhD student in the Center for Neural Informatics, Structures, & Plasticity (CN3), I did a substantial amount of work on NeuroMorpho.Org (Ascoli et al., 2007a). I have actively participated in the increase (from 62,000 to more than 172,000) of the number of neural reconstructions within the database, mastering

different software tools for data processing, literature portal for article mining, and metadata curation. As an active member of NeuroMorpho.Org, I have been a part of, and created different functionalities for end users to analyze and understand the data. Furthermore, I discovered a novel morphometric biomarker that led to the efficient separation of cells of the nervous system. In addition to fostering future research towards understanding the structure-activity-relationship of neurons and glia, as well as the three main classes of glia, I believe my work at NeuroMorpho.Org has already had a substantial impact.

CHAPTER 2: AN OPEN REPOSITORY FOR SINGLE-CELL RECONSTRUCTIONS OF THE BRAIN FOREST

Masood A. Akram¹, Sumit Nanda¹, Patricia Maraver¹, Rubén Armañanzas¹ & Giorgio A. Ascoli¹

¹Center for Neural Informatics, Structures & Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA (USA).

*This chapter was published in *Sci Data* 5, 180006 (2018).

Note: This manuscript, including all the figures, is written by Masood A. Akram.

Abstract

NeuroMorpho.Org was launched in 2006 to provide unhindered access to any and all digital tracings of neuronal morphology that researchers were willing to share freely upon request. Today this database is the largest public inventory of cellular reconstructions in neuroscience with a content of over 80,000 neurons and glia from a representative diversity of animal species, anatomical regions, and experimental methods. Datasets continuously contributed by hundreds of laboratories worldwide are centrally curated, converted into a common non-proprietary format, morphometrically quantified, and annotated with comprehensive metadata. Users download digital reconstructions for a variety of scientific applications including visualization, classification, analysis, and simulations. With more than 1000 peer-reviewed publications describing data stored in or utilizing data retrieved from NeuroMorpho.Org, this ever-growing repository can already be considered a mature resource for neuroscience.

Introduction

At the turn of the millennium the field of neuroinformatics was still in its infancy, but many researchers had already embraced computer simulations as an important approach to ground the understanding of the nervous system into quantitative models (Dayan, 1994; Hines & Carnevale, 2001; Sejnowski et al., 1988). Pioneering work in anatomy and physiology had long established that the tree-like morphologies of neuronal axons and dendrites play fundamental roles in network connectivity and signal processing. Progress in hardware and software had gradually enabled the practice of digitally reconstructing neuronal morphology three-dimensionally from live microscopic feeds or image stacks (Senft, 2011a). These data were especially useful to carry out comprehensive statistical analyses and to build realistic computational models. However, interactions between experimentalists and theoreticians were the exception rather than the norm, and sporadic data exchange relied on serendipitous peer-to-peer contacts.

In that context, NeuroMorpho.Org was envisioned as an open, public-access, online portal where digital reconstructions of neurons from different animal species, brain regions, cell types, experimental conditions, and acquisition methods could be stored and freely retrieved for re-use in any scientific, educational, or artistic application (Ascoli, 2006). The first version of the database was released in 2006 with an initial set of less than 1000 cells (Ascoli et al., 2007a) and was immediately adopted with broad enthusiasm in the neuroscience community (Ascoli, 2007). Since then the repository continuously evolved, adding both new data and improved functionality in each of 30

releases over 11 years. Nevertheless, the key operational principles have remained stable: identifying suitable data in peer-reviewed publications, requesting the source files from authors, and making them available to the community after processing into ready-to-reuse form.

Here we describe the current status and future prospects of this mature resource from multiple perspectives, including those of the users, contributors, and developers. We start with an overview of the content and web-based accessibility, then we describe the community usage of these data, and lastly we provide a backstage view of the curation process enabling dense coverage of this data type (Halavi et al., 2008a).

Results

Database Content

Version 7.3 of NeuroMorpho.Org (fall 2017 release) contains 80,012 cells contributed by 416 laboratories and described in 714 peer-reviewed articles. These data come from ~50 anatomical regions of more than 40 animal species ranging from nematodes to humans and notably encompassing both common experimental model (mouse, rat, fruit fly, etc.) and less usual sources (e.g. manatee, leopard, and giraffe). Most pertinently, all major cell types of the nervous system are represented across a number of dimensions (Fig. 2.1), including sensory receptors, glia, local interneurons and projection neurons releasing excitatory, inhibitory or modulatory neurotransmitters.

For every cell, the database includes the original morphological tracing file as provided by the contributors, the standardized version, and a detailed log of all changes. The basic data format describes the represented trees shape as a set of 3D coordinates (in micrometers), each associated with a tag indicating the type of structure (soma, axon, dendrite, apical, unspecified neurite, glial process or custom-defined), thickness, and the identity of the connected point in the path to the root. In addition to this digital reconstruction file, every cell is displayed with an image, an animation, and a user-interactive in-browser display. For certain users, this latter functionality may require minor JAVA setting adjustments (detailed in the Frequently Asked Questions) due to recent security updates on several operating systems. Moreover, all cells are comprehensively annotated with metadata (Parekh et al., 2015a) regarding the animal subject (species, strain or genotype, sex, age, weight), the cell studied (anatomical region, molecular expression, physiological features), the methodological procedure (experimental condition, tissue sectioning, specimen staining, imaging resolution, tracing system), and linked identifiers (PMID and DOI) of the corresponding referenced publication. Last but not least, a battery of geometric features (length, angles, branch topology, fractal dimension, etc.) is extracted and provided with each individual cell (Scorcioni et al., 2008b).

Web Functionality

NeuroMorpho.Org provides a user-friendly graphical interface to access the data through any modern web browser. Visitors can sample a random set of neurons or browse the entire repository by cell type, brain region, animal species, or contributing laboratory,

corresponding to the intuitive elements most immediately associated with every study (what, where, which, who). In all cases, data can be selected and downloaded or simply dynamically visualized in rapid sequence with simple cursor movements. Furthermore, the separate Neuron Atlas application, which can be freely downloaded from the Browse menu, repurposes an earlier implementation of the Allen Brain Atlas software to map the spatial distribution of neurons across the major anatomical subdivisions of the rodent brain. The interactive 3D display links to the individual neuron pages and also enables direct visualization of the metadata and the single neuronal morphologies (Fig. 2.2).

The web front-end also offers multiple search functions. The metadata search option enables one to quickly identify content by conveniently combining filters through logically organized drop-down menus. For example, selecting mouse or drosophila as species populates the strain menu with completely different content. Choosing ‘C57BL/6’ from the former while setting experimental protocol to ‘in vitro’ pulls (as of version 7.3, released in November 2017) 4803 cells, which can be displayed as a summary or downloaded in bulk. The morphometry search is similarly organized allowing users to

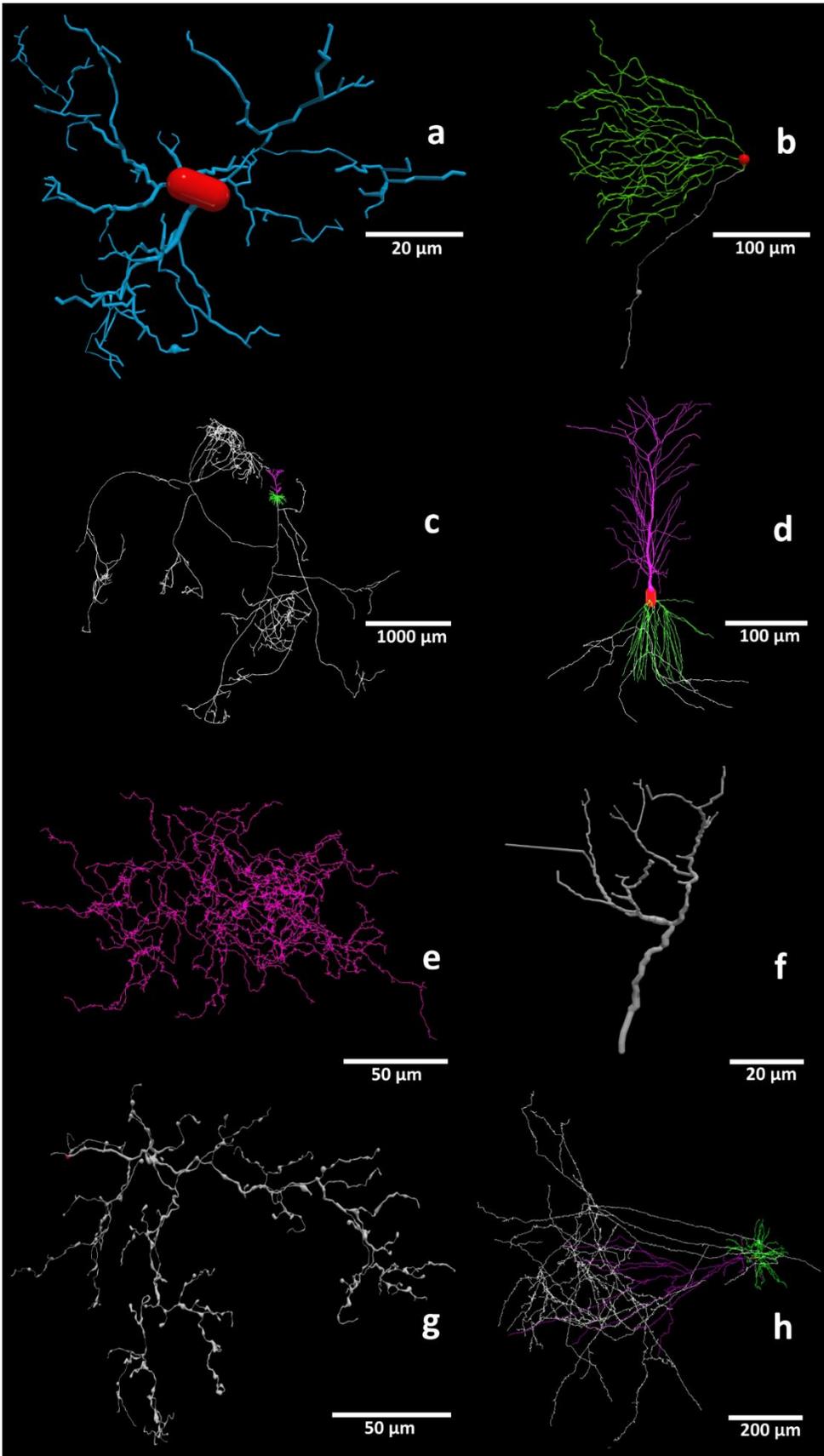


Figure 2. 1. Digital reconstructions of cellular morphologies in NeuroMorpho.Org. (a) Microglia cell from mouse spinal cord (Ohgomori et al., 2016a), with processes in blue and cell body in red. (b) Ganglion cell from mouse retina (Badea & Nathans, 2011), with dendrites in green, axon in silver, and soma in red. (c) Pyramidal neuron from mouse neocortex (Economo et al., 2016), with apical dendrites in magenta, basal dendrites in green, and long projection axon in silver. (d) Pyramidal neuron from rat hippocampus (Ishizuka et al., 1995), with apical dendrites in magenta, basal dendrites in green, and (incomplete) axon in silver. (e) Interneuron from mouse retina (Helmstaedter et al., 2013a), with unspecified neurites in pink. (f) Direction sensitive mechanoreceptor from cricket peripheral nervous system, with unspecified neurites in pink (Jacobs & Theunissen, 2000). (g) Olivocerebellar neuron from rat myelencephalon, with axon in silver (Brown et al., 2012). (h) Pyramidal neuron from cat neocortex (Martinez et al., 2005), with apical dendrites in magenta, basal dendrites in green, and long projection axon in silver.

NeuroMorpho.Org
Version 7.3 - Released: 09/11/2017 - Content: 80008 neurons

HOME BROWSE SEARCH LITERATURE COVERAGE TERMS OF USE HELP

By archive

a

Search: dia
Filter: DIADEM

b

Neuron Type: Interneuron
Archive: Helmstaedter
Neuron: 2003-02-24-A2-L23-dendax

Reference Article
PubMed/Abstract Link DOI Link

Brain Region

Brain Region	A	N
Basic cell groups and regions (Brain)	<input type="checkbox"/>	<input type="checkbox"/>
Cerebrum (CH)	<input type="checkbox"/>	<input type="checkbox"/>
Cerebral cortex (CTX)	<input type="checkbox"/>	<input type="checkbox"/>
Cortical plate (CTXpl)	<input type="checkbox"/>	<input type="checkbox"/>
Olfactory areas (OLF)	<input type="checkbox"/>	<input type="checkbox"/>
Main olfactory bulb (MOB)	<input type="checkbox"/>	<input type="checkbox"/>
Anterior olfactory nucleus (AO)	<input type="checkbox"/>	<input type="checkbox"/>
Hippocampal formation (HPF)	<input type="checkbox"/>	<input type="checkbox"/>
Hippocampal region (HIP)	<input type="checkbox"/>	<input type="checkbox"/>
Ammon's Horn (CA)	<input type="checkbox"/>	<input type="checkbox"/>
Dentate gyrus (DG)	<input type="checkbox"/>	<input type="checkbox"/>
Retrohippocampal region (RHf)	<input type="checkbox"/>	<input type="checkbox"/>
Subiculum (SUB)	<input type="checkbox"/>	<input type="checkbox"/>
Cerebral nuclei (CNU)	<input type="checkbox"/>	<input type="checkbox"/>
Striatum (STR)	<input type="checkbox"/>	<input type="checkbox"/>
Nucleus accumbens (ACB)	<input type="checkbox"/>	<input type="checkbox"/>
Brain stem (BS)	<input type="checkbox"/>	<input type="checkbox"/>
Interbrain (IB)	<input type="checkbox"/>	<input type="checkbox"/>
Thalamus (TH)	<input type="checkbox"/>	<input type="checkbox"/>
Zona incerta (ZI)	<input type="checkbox"/>	<input type="checkbox"/>
Midbrain (MB)	<input type="checkbox"/>	<input type="checkbox"/>
Substantia nigra reticular part (SNr)	<input type="checkbox"/>	<input type="checkbox"/>
Hindbrain (HB)	<input type="checkbox"/>	<input type="checkbox"/>
Medulla (MY)	<input type="checkbox"/>	<input type="checkbox"/>
Hypoglossal nucleus (XII)	<input type="checkbox"/>	<input type="checkbox"/>
Cerebellum (CB)	<input type="checkbox"/>	<input type="checkbox"/>
Cerebellar nuclei (CBN)	<input type="checkbox"/>	<input type="checkbox"/>

Figure 2. 2. Browse functionality. (a) Summary view of selected neurons from the Browse-by-Archive pane, with links to individual entries and options for download. (b) NeuronAtlas, a free downloadable application for exploring digital reconstructions of neuronal morphologies from rat and mouse brains.

select the quantitative ranges of geometric parameters of interest. For instance, querying for neurons with at least 100 branches and a maximum path distance from the soma greater than 1 mm retrieves 5706 cells (v.7.3).

A distinct search mechanism adopts a familiar “Google-like” keyword bar with Boolean logic and wildcards. Alternatively, the advanced OntoSearch function leverages logical reasoning by generalization and specification of hierarchically organized knowledge (Polavaram & Ascoli, 2017a). Moreover, users can query the database content using the Literature Search,

0: "neuron_id"	a	0: "principal cell"	b
1: "neuron_name"		1: "interneuron"	
2: "archive"		2: "projection"	
3: "reference_pmid"		3: "dLGN-projecting"	
4: "reference_doi"		4: "pyramidal"	
5: "note"		5: "neurogliaform"	
6: "age_classification"		6: "granule"	
7: "brain_region"		7: "Embryo-born"	
8: "cell_type"		8: "day4-born"	
9: "species"		9: "sensory receptor"	
10: "strain"		10: "mechanosensory"	
11: "scientific_name"		11: "glutamatergic"	
12: "stain"		12: "GABAergic"	
13: "experiment_condition"		13: "cholinergic"	
14: "protocol"		14: "nitroergic"	
15: "slicing_direction"		15: "serotonergic"	
16: "reconstruction_software"		16: "fast-spiking"	
17: "objective_type"		17: "irregular spiking"	
18: "original_format"		18: "Calretinin (CR)-positive"	
19: "domain"		19: "Somatostatin (SOM)-positive"	
20: "attributes"		20: "Parvalbumin (PV)-negative"	
neuron_name: "AM48-2"	c	pmid: 25838038	d
neuron_id: 2245		doi: "10.1093/cercor/bhv053"	
depth: 36.99		journal: "Cerebral Cortex"	
height: 583.85		species: "Rat"	
bif_ampl_local: 73.53		paper_ID: 1592	
branch_Order: 19		brain_Region: "Neocortex"	
contraction: 0.77		cell_type: "Projection neuron"	
eucDistance: 550.73		data_Status: "Positive response"	
fractal_Dim: 1.05		first_Author: "Narayanan R.T."	
fragmentation: 7096		last_Author: "Oberlaender M."	
n_bifs: 142		#neurons: 153	
n_branch: 285		ocdate: "2015-06-30"	
n_stems: 1		paper_Title: "Beyond Columnar Organization: Cell Type- and Target Layer-Specific Principles of Horizontal Axon Projection Patterns in Rat Vibrissal Cortex"	
partition_asymmetry: 0.56		Details: "Positive response"	
pathDistance: 1015.12		system_Used: "NeuroLucida"	
pk_classic: 2		year: 2015	
surface: 9475.06			
volume: 568.5			
width: 396.67			
diameter: 0.24			
length: 12566.7			

Figure 2. 3. NeuroMorpho.Org’s API gives developers and data miners access to database content and information allowing integration with other neuroinformatics tools. Complete usage documentation and definitions are available at neuromorpho.org/apiReference.html. (a) Sample of available JSON objects characterizing experimental methods. (b) Selected entries specifying the dimension of cell types. (c) Subset of a given cell’s morphometric characteristics defined as attribute-value pairs. (d) Bibliographic details for a data set from an individual publication.

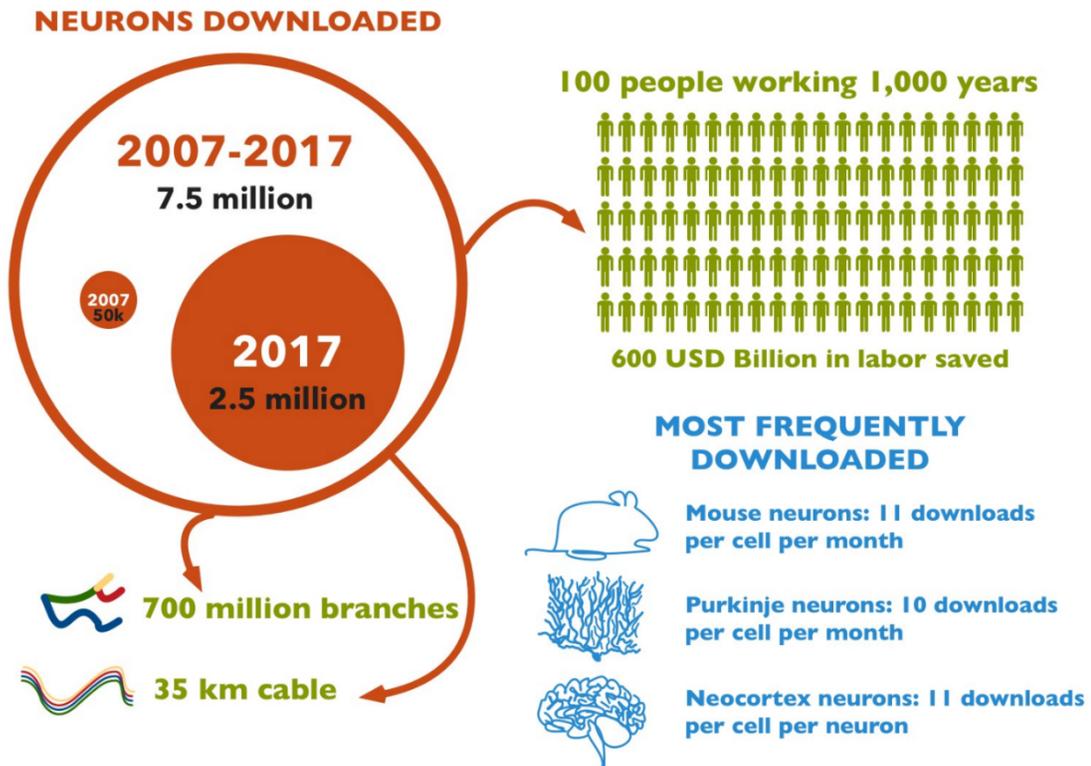


Figure 2. 4. Infographics summary of data download activity from the NeuroMorpho.Org portal.

which is further elaborated on in the Methods section. The most recently added search functionality makes information machine-accessible through a public Application Programming Interface (API) using JavaScript Object Notation (JSON), a lightweight data-interchange format. The NeuroMorpho.Org API provides large-scale data access through anchors to neuron metadata, morphometric measurements, and literature references, in addition to brain regions, animal species, and more (Fig. 2.3).

Usage Statistics

Digital reconstructions shared through NeuroMorpho.Org are re-used for various purposes including training and dissemination, implementation of morphologically realistic computational simulations, assessments of potential synaptic contacts, validation of developmental growth models, and comparative statistical analyses across different neuron types, anatomical locations or experimental conditions. The usage of the repository can be quantified in terms of unique website visits, number of downloads, and derived publications. As of December 2017, the web page has been visited above 270,000 times from more than 75,000 unique internet addresses in 166 countries. The number of hits increased from less than 3000 in the first year to over 15,000 in the June-August 2017 trimester. Notably, two-thirds of this traffic comes from the United States and China.

A total number of 7.5 million neurons were downloaded in standardized format along with 4.5 million ancillary files (original “as received” morphologies and standardization logs) before the 7.3 release. Approximately 50,000 files were

downloaded in the first year of the project up to a record 2.5 million in June-August 2017 (Fig. 2.4). The download volume can also be tallied in number of neurite branches (700 million) or total cable length (35 km). Based on an average of three working days per neuron (one for histology, one for imaging, one for tracing), it would take nearly 100,000 people working for a year to produce the amount of data downloaded in the course of this project. The most popular species and brain region are respectively mouse and neocortex (11 downloads per cell per month). The most frequently downloaded neuron types are Purkinje, newborn granule, and extraverted pyramidal cells (10 downloads per cell per month). Complete statistics are updated online at every release.

The standard way to quantify impact in scientific research is by tracking citations. At least 543 publications cite NeuroMorpho.Org, 353 of which describing results that were directly based on data obtained from this resource. This “secondary” usage substantially adds to the primary service of providing a data repository and standardization process. In this regard, before describing representative examples of re-usage application in the next section, it is worth mentioning that leading publishers, including Springer-Nature, Elsevier, and the Public Library of Science, recommend NeuroMorpho.Org as a trusted database for deposition of digitized morphological reconstructions. This is particularly important given the growing opportunity for ‘data publication’ (Akins et al., 2017).

Representative Applications

One of the most typical usages of digital reconstructions of neuronal morphology consists of constraining and validating data-driven models of electrophysiological activity or of developmental structural dynamics. For example, pyramidal cells taken from NeuroMorpho.Org are frequently studied in simulations of activity (Popovic et al., 2015) or structure (López-Cruz et al., 2011a). Often these studies are extended to obtain novel computational or theoretical conclusions. For instance, the comparative analysis of dendritic complexity across animal taxonomy was recently related to the discriminative capacity of signal integration (Zippo & Biella, 2015), and the virtual generation of neuronal trees revealed a balance between the minimization of total wiring and signal conduction time (Cuntz et al., 2010). An open problem increasingly addressed with morphological quantification is the unbiased classification of neuron types (Armañanzas & Ascoli, 2015a). Application of co-clustering techniques on a large set of neocortical cells from many different labs and experimental preparations revealed a clear separation between principal neurons located in different cortical areas and depths (Y. Lu et al., 2015a). Leveraging modern machine learning methods also yielded methods for content-based retrieval of morphologies, where similar neurons can be efficiently identified from a large collection based on sparse examples (Conjeti et al., 2016a).

Earlier examples of secondary discoveries have been extensively reviewed elsewhere (Parekh & Ascoli, 2015a). Evidence of broader impact has also emerged beyond basic scientific research. The open availability of digital data is encouraging uptake in divulgation efforts, from undergraduate education (Chu et al., 2015a) to 3D

printable models (McDougal & Shepherd, 2015). By their own nature, many expository or outreach applications are not associated with peer-reviewed references, but are nonetheless beneficial to global societal progress. As representative examples we mention the selection as “Site of the Month” in Neuroscience For Kids (faculty.washington.edu/chudler/neurok.html), the creation of an online teaching module in BrainU (brainu.org), and the use as testing dataset in a past competition of the Chinese Applied Math Olympiads (neuromorpho.org/china_contest.jsp). Similarly, data from NeuroMorpho.Org are used to train commercial expert systems (drvtechnologies.com/aivia5) and to produce multimedia art installations (vimeo.com/191338612). Only a fraction of these outcomes could be predicted at the outset, reinforcing the notion that open data availability may generate genuinely new opportunities for discovery and creativity that would be missed in a more restrictive “sharing upon request” model.

Discussion

After more than a decade of continuous operation, NeuroMorpho.Org is now considered a stable and mature resource in the neuroscience community. Strikingly, we curators receive as much positive feedback and expressions of gratitude from data contributors as from data users. On the one hand this reflects the value added by the systematic standardization and annotation processes that are detailed in the Methods section. On the other, the above observation strongly refutes the entrenched view of a unidirectional cost-benefit flow from the “heroic” experimental data producers to the

“parasitic” computational data modelers (Longo & Drazen, 2016). On the contrary, it corroborates the alternative idea of synergistic cooperation even in the absence of direct collaboration, since re-usage demonstrably augments the impact of the original dataset (Ascoli et al., 2017c). At the same time, we treasure constructive criticisms from the community, and the vast majority of new and enhanced functionalities were implemented over time to address outside suggestions and requests.

The pace of growth in database content continued to rise substantially over the lifespan of the project. While reconstructions initially accumulated at an average rate of less than 500 per year, most recently the yearly increase of data received passed the figure of 25,000. This rather dramatic acceleration is due to several compounded causes. Most prominently, the field of computational neuroanatomy became increasingly ‘hot’ and the number of publications describing digital reconstruction of neuronal and glial morphology has grown by an order of magnitude over the past 11 years (from ~7 to ~70 articles per month), even faster than the general upward trend in biomedical research in general and neuroscience in particular. Secondly, the progressive automation of the tracing process is boosting the typical dataset size (Nanda et al., 2015), from less than 20 neurons per study in the first year of operation to over 100 in the most recent one, with a notable single contribution of over 16,000 reconstructions (Chiang et al., 2011). Third, the attitude towards data sharing is gradually improving, from a hesitant 25% of positive responses in 2006 to a more encouraging 55% in 2017. In order to keep up with the expanding volume of incoming datasets, the information technology and data

management infrastructure had to be progressively modernized and improved, as described in the Methods.

The content is also adapting to the parallel evolution of neuroscience research. For instance, glia (astrocytes, oligodendrocytes, microglia, and other types), once considered the “dark-matter” of the nervous system or the “forgotten” brain cells, are now recognized to contribute crucially to fundamental physiological processes, such as neural development (Zuchero & Barres, 2015), signal processing (I. Song & Dityatev, 2018), and synaptic plasticity (Dzyubenko et al., 2016). Furthermore, glial cells are involved in major neuropathologies, including Alzheimer’s (J. Rodríguez et al., 2016) and Parkinson’s (Greggio et al., 2017) diseases, stroke (F.-E. Song et al., 2017) and epilepsy (Vezzani et al., 2017), as well as traumatic brain (Sajja et al., 2016) and spinal cord injuries (Ahuja et al., 2017). Specifically, glia morphology represents one of the most useful biomarkers of brain function and dysfunction, as exemplified by enlargement of activated microglia upon rising neuronal death (Plaza-Zabala et al., 2017), loss of myelination related to retraction of oligodendrocytes (Dulamea, 2017), and altered astrocyte architecture in response to pharmacological treatment (Ahmed et al., 2017) and neurotoxicity (Blanco-Suárez et al., 2017). In line with this mounting emphasis and awareness, we started receiving spontaneous requests from researchers to include digital tracings of glia arbor morphology into the repository. We thus appropriately modified the metadata schema, search engine, and ingestion scripts, and release 7.3 now contains 3069 glial reconstructions from 8 initial datasets.

In the foreseeable future, NeuroMorpho.Org is similarly expected to keep growing both quantitatively (content amount) and qualitatively (content type). Continuous integration with community resources will emphasize neuronal classification both through comprehensive programs such as the National Institutes of Health-fostered

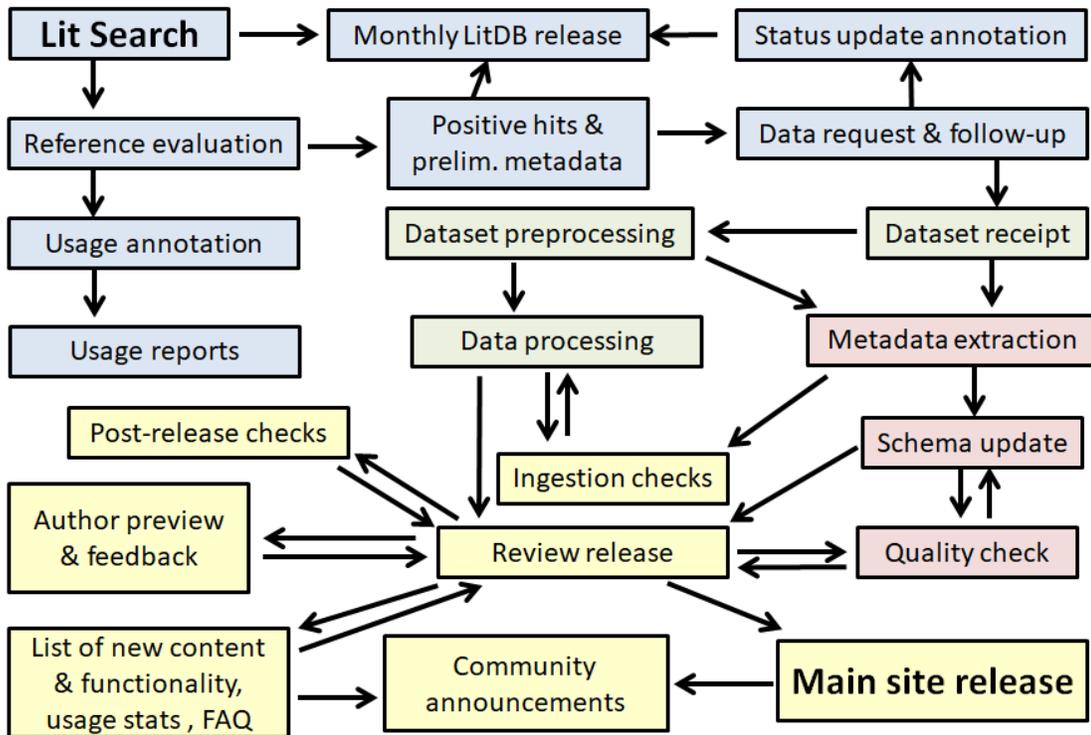


Figure 2. 5. Flowchart of the operational procedures of NeuroMorpho.Org from start (Literature search) to end (Main site release), highlighting literature mining (blue), morphological standardization (green), metadata annotation (pink), and data release (yellow).

Brain Initiative Cell Census Network (BICCN) and domain-specific projects like Hippocampome.Org (Wheeler et al., 2015). This growth is going to necessitate the design and implementation of ergonomic, smart, and robust tools for richer metadata annotation benefiting authors and data curators alike. Moreover, it will be essential to incorporate the anatomical embedding of the reconstructed neurons in a representation of the surrounding tissue. Technically, this only requires the specification of a triad of non-aligned three-dimensional neuronal tracing points relative to a common coordinate framework. This augmented representation is within reach at least for fly, mouse, and human datasets, and would synergistically complement ongoing efforts to capture correlated pre- and post-synaptic circuit connectivity along with the digital morphology (Takemura et al., 2017). Last but not least, progress is advancing towards the extension of the arbor reconstruction format to accommodate temporal and multi-channel information (Nanda et al., 2018b). These additions are especially important in light of recent experimental breakthrough in live time-lapse super-resolution imaging which could reveal critical biochemical details of neuronal dynamics.

Methods

This section describes the current operational pipeline of NeuroMorpho.Org. The overall process can be broadly divided into four main task aggregates: literature mining, data processing, metadata annotation, and public release (Fig. 2.5). The execution of most steps is logically serial for each dataset, though the team organization is largely parallel:

while releases are lumped into a few versions per year, new datasets are continuously identified, requested, received, processed, and annotated.

Given NeuroMorpho.Org's mission of freely distributing digital reconstructions from peer-reviewed publications, it is natural for the data identification process to begin with literature mining (Halavi et al., 2012a). Using a comprehensive battery of keywords in appropriate combinations and several full-text search engines, we collect a set of newly published articles every month which possibly describe tracing datasets. Since the monthly queries were optimized to minimize the number of missed articles, only approximately a third of the shortlisted potential hits actually contain relevant data. Furthermore, users can also directly suggest articles to mine by checking the reference status using the Literature Search function. Team curators carefully evaluate each article and extract preliminary metadata information for every positive instance, including number of reconstructions, brain region, cell type, reconstruction system, bibliographic reference, and corresponding author contact. These minimal details are necessary and sufficient to invite data submission. The outcome of the request (after multiple reminders and follow-ups as needed) determines whether each given dataset is available (received and in processing pipeline) or unavailable (unanswered: ~70%, declined: ~15%, or declared lost: ~15%). A small but increasing proportion of data (5% in 2017) is sent spontaneously by the authors before publication. In addition to new data, the literature mining staff also tracks published usage of data downloaded from the repository. The

entirety of the above information is recorded in a public literature database (Ascoli, 2015c) that is updated monthly (Fig. 2.6).

Availability Status ^a		Summarized by Reconstruction Information ^b	
PMID/DOI	Contacted authors	Human	Monkey
28060701	Perisomatic GABAergic synapses of basket cell effectively control principal neuron activity in amygdala networks Requested on J. M. Veres , G. A. Nagy, N. Hajos 02/17/2017 eLife In processing pipeline	17924570	111 Pyramidal neurons using NeuroLucida
27353294	Early-onset Purkinje cell dysfunction underlies Cerebellar ataxia in proximal multifunctional protein-2 deficiency Requested on Stephanie De Munter , Simon Verheijden, Esther Vanderstuyft, Ana Rita Malheiro, Pedro Brites, David Gall, Serge N. Schifmann, Myriam Baes 08/10/2016 Neurobiology of Disease In repository	24453325	15 GABAergic inhibitory interneurons using NeuroMantic
26615780	DNA repair factor BRCA1 depletion occurs in Alzheimer brains and impairs cognitive function in mice Requested on Eisa Suberbielle , Biljana Dhukic, Mark Evans, Daniel H. Kim, Praveen Taneja, Xin Wang, Mariel Finucane, Joseph Knox, Kaitlyn Ho, Nino Davidze, Eliezer Maslah, Lennart Mucke 12/31/2015 Nature Communications No response	22002739	291 Subplate neurons using NeuroLucida
24277868	Mechanisms by which a CACNA1H mutation in epilepsy patients increases seizure susceptibility Requested on V. S. Ecker , A. Shcheglovitov , D. Dey, C. C. Yap, B. Winckler, E. Perez-Reyes 04/07/2014 The Journal of Physiology Request declined	20187150	282 Type I, Type II and Pyramidal neurons using NeuroLucida
27057368	Nutritional Omega-3 Deficiency Alters Glucocorticoid Receptor-Signaling Pathway and Neuronal Morphology in Regionally Distinct Brain Structures Associated with Emotional Deficits Requested on T. Larnieu , M. L. Hilal, Smedt-Peyrusse. De, N. Sans, S. Laye 02/24/2017 Neural Plasticity Positive response	24920624	54 Interneurons using Simple Neurite Tracer
25546402	Type 2 wide-field amacrine cells in TH::GFP mice show a homogenous synapse distribution and contact small ganglion cells Requested on B Bruggen , A Meyer, F Boven, R Weiler, K Dedek 08/01/2017 European Journal of Neuroscience Invited	14627644	113 Pyramidal neurons using NeuroLucida
		24878822	405 Medium spiny neurons using Imaris
		21871563	768 Pyramidal and Medium spiny neurons using NeuroLucida
		27173096	50 Pyramidal Neurons and Interneurons using Neurite Tracer
		22162008	119 Principal neurons using NeuroLucida
		25838038	153 Projection neurons using NeuroLucida
		27238893	16 Medium spiny neurons using NeuroLucida

Figure 2. 6. Literature database and search. (a) Retrievable information by availability status. To avoid singling out individual authors, the identifiers for declined or unanswered queries are fictional in this illustration; actual unabridged data are publicly posted online in the ‘Literature Coverage’ section of NeuroMorpho.Org. (b) Preliminary metadata annotation for all references deemed to describe digital reconstructions of neuronal morphology.

If the authors choose to share their traced cells, the next steps are metadata annotation and morphological standardization. The authors are invited to contribute to the

metadata annotation; their input is complemented with information extracted from the corresponding publication and team curators formalize the resulting description using control vocabularies and formal ontologies prior to ingestion (Polavaram & Ascoli, 2017a). New terms are judiciously added as needed to the corresponding schema. Meanwhile, every neuron is assigned a unique machine-readable identifier while also maintaining its original human-assigned name after minor modifications if necessary to avoid occasional duplications or special characters known to interfere with smooth web interactions.

The raw neuronal reconstructions, which can be acquired in more than 20 different tracing systems and file types, are then translated into a common non-proprietary format (Cannon et al., 1998a) and undergo a series of consistency checks such as detection of non-positive diameters, overlapping points, and disconnected branches among many others. A comprehensive description of the complete standardization process is also posted on NeuroMorpho.Org along with the open-source checking software. Inconsistencies are edited using a variety of programs, including CVapp (Cannon et al., 1998a), Neuromantic (Myatt et al., 2012), neuTube (Feng et al., 2015), Vaa3D (Peng et al., 2011), TREES Toolbox (Cuntz et al., 2011), and several more as needed (Parekh & Ascoli, 2013). Importantly, all edits are tracked in accompanying log reports.

After completing morphological editing, data processing is finalized by extraction of morphometric parameter and generation of static images (Torben-Nielsen, 2014a) and rotating animations for each neuron. All data, metadata, and ancillary files are then ingested into the previous version of the database to produce a password-protected Review site. This step is essential to enable all contributing authors to preview their data and request changes prior to public release. Furthermore, at this stage we can implement a number of cross-checks, which include: detecting duplicated cells, broken web links, and missing data files; ensuring the accuracy and consistency of new metadata entries and proper integration of enhanced functionalities; and updating usage statistics and frequently asked questions. Upon resolving any ensuing issues if any, the new version is finally openly released and publicly announced through mailing lists and social media.

As with any large-scale scientific database, occasional errors in data or metadata are detected and reported after public release. We try and correct all mistakes immediately if possible, without waiting for the next release. In case of simple typos, the correction is implemented silently. Any change in metadata, in contrast, is disclosed in the following release with a link to detailed documentation (see e.g. last bullet of v6.3 in the “What’s New” page). When correcting the morphological files themselves, a link to the previous version is provided in the Notes of the corresponding neuron page (cf. sixth bullet of v6.1 in the “What’s New” page).

Author's contributions

All authors are core members of the NeuroMorpho.Org team. MAA is responsible for literature evaluation and metadata annotation; SN is responsible for data conversion and external tool integration; PM is responsible for literature management and access statistics; RA is the technical lead, supervises the software programmers, and is responsible for data ingestion/release as well as database maintenance/development; GAA is the project founder and overall director, interacts with data owners, and is responsible for data requests, management, processing, and quality-control. MAA and GAA wrote the manuscript and created the illustrations with the remaining co-authors providing edits and crucial feedbacks.

Competing interests

The author(s) declare no competing financial interests.

Acknowledgements

The authors thank Zulfa Ghannouchi for assistance in literature mining, Kayvan Bijari for useful discussion, Usama Saifi for help with the illustrations, Pruthvik Narayanaswamy and Suyameendra Wadki for software programming and data ingestion, and the undergraduate data processing team (neuromorpho.org/acknowl.jsp#NMO) for careful morphological proofing. This work was supported in part by grants R01NS039600 (NIH), R01NS086082 (NIH), U01MH114829 (NIH), and DBI-1546335

(NSF). Publication of this article was funded in part by the George Mason University Libraries Open Access Publishing Fund.

CHAPTER 3: EFFICIENT METADATA MINING OF WEB-ACCESSIBLE NEURAL MORPHOLOGIES

Masood A. Akram, Bengt Ljungquist, Giorgio A. Ascoli

Center for Neural Informatics, Structures & Plasticity, Krasnow Institute for Advanced Study, and Department of Bioengineering, Volgenau School of Engineering, George Mason University, Fairfax, 22030, VA, USA

*This chapter was published in *Progress in Biophysics and Molecular Biology*, 168, 94–102 (2022).

The authors declare no conflict of interest

Keywords: NeuroMorpho.Org, digital reconstructions, neuron, glia, morphometry, application programming interface, summary reporting.

Note: The code for summary reporting was written by the developers of NeuroMorpho.Org. The testing, utilization, statistical analysis, and manuscript writing was done by Masood A. Akram.

Abstract

Advancements in neuroscience research have led to steadily accelerating data production and sharing. The online community repository of neural reconstructions NeuroMorpho.Org grew from fewer than 1,000 digitally traced neurons in 2006 to more than 140,000 cells today, including glia that now constitute 10.1% of the content. Every reconstruction consists of a detailed 3D representation of branch geometry and connectivity in a standardized format, from which a collection of morphometric features is extracted and stored. Moreover, each entry in the database is accompanied by rich

metadata annotation describing the animal subject, anatomy, and experimental details. The rapid expansion of this resource in the past decade was accompanied by a parallel rise in the complexity of the available information, creating both opportunities and challenges for knowledge mining. Here, we introduce a new *summary reporting* functionality, allowing NeuroMorpho.Org users to efficiently download digests of metadata and morphometrics from multiple groups of similar cells for further analysis. We demonstrate the capabilities of the tool for both glia and neurons and present an illustrative statistical analysis of the resulting data.

Introduction

Combining microscopes with computers in neuroscience has enabled routine reconstruction of complex branching structures, like dendrites, axons, and glial processes, into digital files, thus describing morphological information in dedicated data files (Parekh & Ascoli, 2015b; Senft, 2011b). For instance, the swc format (Cannon et al., 1998b) stores a standardized 3D digital representation by defining a tree structure of nodes, their coordinates, and radii. The largest public inventory of such standardized cellular reconstructions, NeuroMorpho.Org, was launched in 2006 with an initial data set of fewer than 1000 cells (Ascoli et al., 2007) to provide unhindered access to all digital tracings of neuronal morphology freely shared by willing researchers upon request. Glial cells were added to this online repository in version 7.1, publicly released in 2017 (Ascoli et al., 2017a). With the latest version 8.0.111 (2021-03-19), glial cells constitute over 10% of more than 145,000 cellular reconstructions. Hundreds of laboratories worldwide

continuously contribute data sets of neurons and glia to NeuroMorpho.Org (Ascoli, 2015b), which then are processed (including conversion into standardized swc format), morphometrically quantified, annotated with comprehensive metadata, and ingested for public access (Akram et al., 2018c).

Morphological studies have benefited from substantial advancement in microscopy and tracing software (Peng et al., 2017b), enabling automation. As a result, the rate with which cells are traced has increased. Together with an improved attitude towards data sharing in neuroscience (Gleeson et al., 2017b), this has enabled the data deposited to NeuroMorpho.Org to grow substantially into a true Big Data resource. This significant increase in the amount of data was unavoidably accompanied by a parallel rise in metadata complexity (**Figure 3.1**). Upon first launching, NeuroMorpho.Org included digital reconstructions from 7 species, 31 neuron types, and 16 brain regions, shared by 21 individual laboratories. The latest major release, 8.0, tallies 76 species, 1183 cell types, 372 brain regions, and 686 contributing labs (Fig. 3.1A-i/ii). On the one hand, community usage also continues to grow, with 22 million downloads and 1.2 million unique accesses worldwide (Fig. 3.1A-iii). On the other, the sheer diversity in important details such as experimental conditions, labeling methods, histological protocols, and tracing software (Fig. 3.1B) has created both opportunities and challenges for knowledge mining.

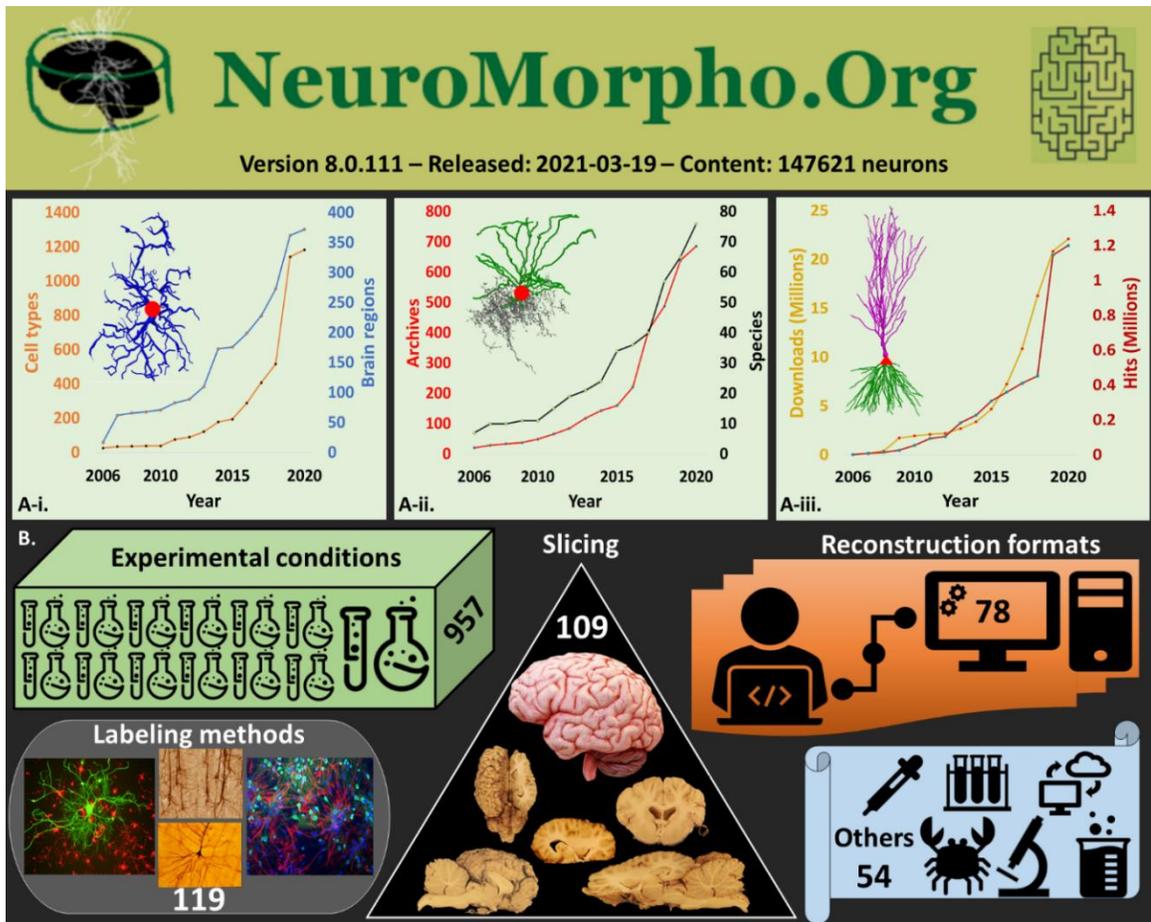


Figure 3. 1. Growth and complexity of data in NeuroMorpho.Org. **A-i.** Numbers of cell types and brain regions added to the database from 2006 to 2021. Inset: representative mouse spinal cord glial cell from the Ohgomori archive (Ohgomori et al., 2016b). **A-ii.** Numbers of contributing labs and represented species since 2006. Inset: mouse neocortical chandelier cell from the Gonzalez-Burgos archive (Miyamae et al., 2017). **A-iii.** Numbers of downloaded cells from and unique online hits to the repository. Inset: guinea pig hippocampal pyramidal cell from the Wittner archive (Wittner & Miles, 2007). **B.** Infographic display of the diversity in experimental conditions, slicing directions, reconstruction formats, labeling methods, and other metadata details in the current release of NeuroMorpho.Org. Images from panel **B** are modified from commons.wikimedia.org under a Creative Commons Attribution-Share Alike 2.5 Generic license.

Here, we present an original tool, *summary reporting*, which provides novel data mining capabilities and allows researchers to generate and download comprehensive metadata and morphometric features for arbitrary user-selected subsets of the database content. Cataloging and grouping cells by homogenous metadata and morphometrics provides an intuitive overview of the database content and facilitates further scientific discovery (Halavi et al., 2012b), including classification (Armañanzas & Ascoli, 2015b), computational modeling (Berzhanskaya et al., 2013; Koene et al., 2009; López-Cruz et al., 2011b), visualization and quantification (Chu et al., 2015b; Wan et al., 2015a), statistical analyses (López-Cruz et al., 2011b; Polavaram et al., 2014a), connectomics estimates (J. Lu, 2011; Ropireddy & Ascoli, 2011), and electrophysiological simulations (Lazarewicz et al., 2002; Vermaas et al., 2020). This article demonstrates the capabilities of the tool, including statistical analysis of the results to illustrate possible usages and applications.

Materials and Methods

Every reconstruction in NeuroMorpho.Org is annotated with **metadata** providing a detailed qualitative description of the cell (Parekh et al., 2015b). This information consists of 35 distinct fields pertaining to the animal subject, anatomy, completeness, experimental conditions, and provenance. Part of this information is often provided directly from the research groups depositing their digital tracings and is then complemented by in-house curators through analysis of the reference peer-reviewed article. Metadata curation used to be a purely manual annotation process, but a user-

friendly online metadata portal was recently introduced that is more reliable and less labor-intensive (Bijari, Akram, et al., 2020c).

Along with the reconstruction files and accompanying metadata, 21 standard **morphometric parameters** are also automatically calculated for each cell (**Figure 3.2**) using the L-Measure tool (Scorcioni et al., 2008c). They include soma surface, number of branches, length, volume, angles, topological asymmetry, fractal dimension, and taper rate among several others. A new characterization of arbor geometry was added in version 7.5 (2018) of NeuroMorpho.Org, called **Persistence Diagram Vectors (PDVs)**. Persistence diagrams compactly describe the branching patterns of neuronal and glial trees using the mathematical formalism of algebraic homology (Kanari et al., 2018a). NeuroMorpho.Org PDVs formally subsume the distributions of the numbers of branches as a function of distance from the soma (“Sholl plots”) into 100-dimensional normalized real number values (Y. Li, Wang, et al., 2017b).

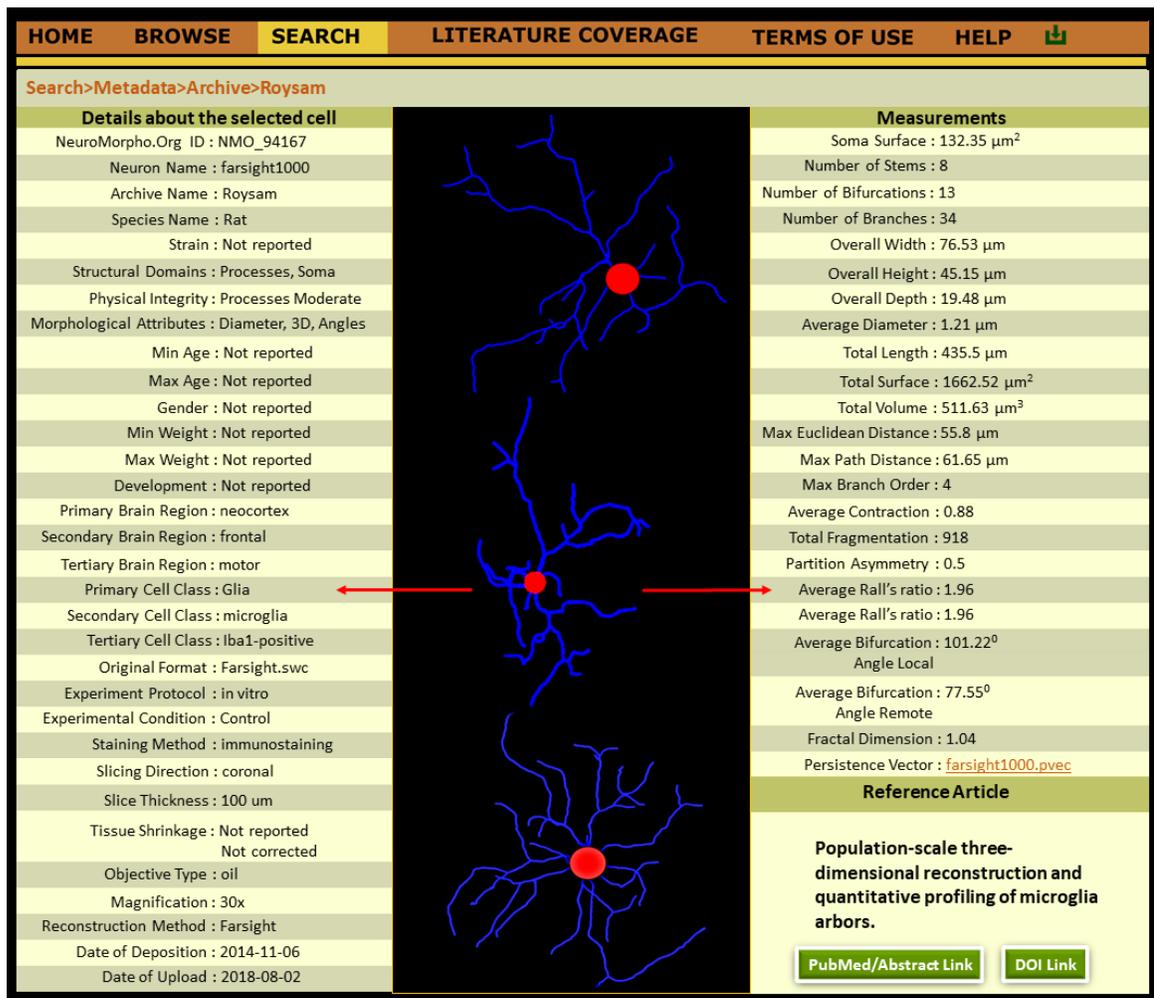


Figure 3. 2. Data, metadata, and morphometry content of NeuroMorpho.Org. **Left:** metadata of a glial cell from the Roysam archive (Megjhani et al., 2015a). The various fields specify for instance the species (rat), cell type (Iba1-positive microglia), brain region (motor cortex), experimental conditions (control), and protocol (in vitro). **Middle:** three digital reconstructions from this group of cells. **Right:** morphometric parameters derived from L-Measure and the bibliographic citation to the reference article for these data.

Summary reporting now enables users to select a subset of data (such as glia) and for that subset produces comma-separated-value (CSV) files of morphometry, persistence diagram vectors, and metadata details for all matching cells grouped in distinct (and internally homogeneous) metadata groups. The implementation of this functionality uses the NeuroMorpho.Org Application Programming Interface (**API**). The new summary reporting tool is divided into three parts: a front-end, a back-end, and a metadata term cache proxy. The front-end, programmed in JavaScript, creates the user interface by fetching metadata terms from the cache proxy and then, as per user interaction, feeds the queries to the back-end. The back-end, developed in Python, fetches the data from the NeuroMorpho.Org API as well as directly from the MySQL database and organizes the output from these data sources. The API is in turn connected to Apache Solr, a data search and index replication software, which maximizes the speed of common queries from the underlying MySQL database. The metadata term proxy, also written in Python, provides a cached copy of all the possible metadata terms that are used for populating the user interface, to make the user experience faster. The code for all three modules is available open-source at github.com/NeuroMorpho/summary-report.

As a representative testbed, we used summary reporting to download the metadata, morphometrics, and persistence diagram vectors of all glial cells (from any species) as well as all monkey principal neurons in the database. This resulted in summaries of 11,569 glial cells divided into 218 distinct groups by their metadata characteristics and 1,129 monkey principal neurons divided into 77 distinct groups. We

thus compared each of those 218 glial groups against all other 217 and each of the 77 monkey principal neuron groups against all other 76 to determine how many of the 35 metadata fields were different across every pair in both cases.

Unlike metadata, the 21 morphometric parameters consist of continuous variables whose numerical value is unique for every cell. We thus applied pairwise two-tailed t-tests, implying the need to correct for multiple comparisons. While 218 groups give rise to 23,653 pairings (218 choose 2), which for 21 morphometric parameters produce a total of 496,713 (23,653 times 21) possible tests, some pairwise comparisons are excluded a priori, because not every morphometric can be extracted for all cells. For example, some datasets may not contain the cell body (yielding a ‘not applicable’ value for soma surface), while others might be traced in 2D from a maximum intensity projection (yielding a ‘not applicable’ value for arbor depth). After discounting excluded comparisons, the number of valid tests was 396,212. We thus established statistical significance ($\alpha=0.05$) at the threshold p-value of $1.26195 \cdot 10^{-7}$ (0.05 divided by 396,212) after Bonferroni correction for multiple testing. Similarly, we applied pairwise two-tailed t-tests on the monkey principal neuron dataset. The total number of possible tests was in this case 2,926 (77 choose 2), for a total number of possible tests of 61,446 (2,926 times 21). The number of valid tests after excluded comparisons was 37,977 and the threshold p-value was $1.31659 \cdot 10^{-6}$ after Bonferroni correction for multiple testing.

Lastly, for the **PDV comparisons**, we first calculated the dot products of the vectors corresponding to each pair of cells within and between groups. Each dot product produces a pairwise arccosine distance. If two cell groups differ in terms of their branch distribution, it is expected that the distance of two cells belonging to the same group will tend to be smaller than that of two cells belonging to separate groups. Based on this logic, we computed t-tests on the arccosine values by comparing within-group to between-group distances. For example, in the comparison of group 1 with group 2, the arccosine values of all cell pairs within group 1 and of all cell pairs within group 2 were tested against the arccosine values of all pairs with one cell from group 1 and the other from group 2. Since the test is only meaningful if between-group distances are greater than within-group, a one-tailed t-test was applied. The threshold p-value for the glia dataset was set to $2.11390 \cdot 10^{-6}$ (0.05 divided by 23,653 group pairs) after Bonferroni correction for multiple comparisons, and the threshold p-value for the monkey principal neuron dataset was set to $1.709 \cdot 10^{-5}$ (0.05 divided by 2,926 group pairs).

Results

The newly introduced summary reporting tool provides a user-friendly interface to efficiently organize and download metadata, morphometrics, and persistence diagram vectors for any subsets of glia and neurons of interest. These reports may then be further utilized for computational and statistical modeling. The functionality can be accessed

from any page on NeuroMorpho.Org¹ by selecting “Summary reporting” from the ‘Search’ drop-down options of the main top-banner menu (**Figure 3.3**).

The screenshot displays the NeuroMorpho.Org search interface. At the top, a navigation bar includes links for HOME, BROWSE, SEARCH, LITERATURE COVERAGE, TERMS OF USE, and HELP. Below this, a breadcrumb trail shows 'Search > Generate Report'. The main area is filled with various filter categories, each with a dropdown arrow and several selectable options:

- Species:** Mouse, Rat, Drosophila melanogaster, Semipalmated sandpiper, Rabbit
- Gender:** Male, Female, Male/Female
- Minimum Weight:** (empty)
- Maximum Weight:** (empty)
- Development:** Young, Adult, Old
- Minimum Age:** (empty)
- Maximum Age:** (empty)
- Brain Region Level 1:** Hippocampus, Neocortex
- Brain Region Level 2:** (empty)
- Brain Region Level 3:** (empty)
- Cell Type Level 1:** Glia
- Cell Type Level 2:** Microglia, Oligodendrocyte
- Cell Type Level 3:** (empty)
- Structural Domain:** (empty)
- Physical Integrity:** (empty)
- Morphological Attributes:** (empty)
- Protocol:** In vitro, In vivo, Culture
- Experiment Condition:** Control, High fat diet, Traumatic brain injury
- Stain:** Immunostaining, Golgi
- Slicing Thickness:** (empty)
- Slicing Direction:** Coronal, Parasagittal, Horizontal
- Reconstruction Software:** (empty)
- Objective Type:** Oil, Water, Dry, Electron microscopy
- Objective Magnification:** (empty)
- Archive:** Diniz, La Barbera, Weil, Bilkei-Gorzo
- PMID:** (empty)
- Original Format:** (empty)
- Date of Deposition:** (empty)
- Date of Upload:** (empty)

At the bottom of the filter area, a summary box indicates '508 glia found'. It features a 'View' dropdown menu currently set to 'Type of Data', which has opened to show options: Groups, Morphometrics, Persistence Vectors, and All. To the right of the dropdown is a 'Download' button. Below the summary box are two buttons: 'Reset' with a circular arrow icon and 'Hits' with a right-pointing arrow icon.

Figure 3. 3. Summary reporting graphical user interface. An intuitive set of drop-down menus corresponding to the annotated metadata dimensions enables users to set the selection filters for choosing the targeted subset of data. The entries represented here serve the sole purpose of illustrating the diversity of possibilities and the tool functionality rather than constituting scientifically meaningful filtering choices. After selecting the desired entries, the webpage allows users to determine the selected data size (“Hits”) and to download CSV files of metadata grouping, morphometric parameters, and persistent diagram vectors.

Users can select specific entries across various metadata fields to filter the desired target content. Multiple selections from distinct fields are interpreted as linked with the Boolean operator AND. For example, selecting ‘mouse’ from the species menu and ‘glia’ from the primary cell type menu will identify all and only mouse glia reconstructions in the database. In contrast, multiple selections from the same field are linked with OR logic. For example, selecting ‘hippocampus’ and ‘spinal cord’ from the primary brain region menu will identify any entry from either of those two anatomical areas. Combining all four of the above filters would identify all glial cells from mouse hippocampus or spinal cord. The “hits” button at the bottom of the page returns the number of cells shortlisted based on the specified details, followed by the option of downloading three reports respectively for metadata, morphometrics, and persistence diagram vectors, individually or collectively. The metadata report lists the specific annotation entries in each of the 35 dimensions for every distinct group of cells within the filtered subset, along with the number of cells in that group. The morphometric report lists the individual values of each of the 21 variables for every single cell in the subset along with the group belonging of that cell. Lastly, the PDV report lists the 100 scalar

components of the vectors corresponding to every single cell in the subset, again together with the group belonging of that cell.

By providing a framework for filtering subsets of data by combinations of metadata annotations (**Fig. 3.3**), summary reporting thus collates the respective information from the archives in the database, creating the CSV files dynamically. In the downloadable CSV files, all cells meeting the selected criteria are partitioned into homogeneous metadata groups whereas cells from distinct groups differ in at least one metadata detail. For instance, in the above example identifying all glial cells from mouse hippocampus or spinal cord, hippocampal cells will be in separate groups from spinal ones; moreover, hippocampal glia will be further separated into distinct groups based on any difference in mouse strain, developmental stage, slicing orientation, labeling method, or other experimental details. Such data extraction and organization steps occur in just seconds even for considerably large data sets, thereby keeping user wait time to a minimum. For instance, it takes less than 2 seconds to generate and download the reports for 3,055 astrocytes (80 groups). **Table 3.1** provides additional benchmarks for representative user filter selections.

Table 3. 1. Summary reporting use case scenarios. Wall clock time to shortlist and download the CSV files for representative user selections of different metadata combinations along with the total number of hits, the corresponding count of distinct cell groups, and the resulting file size.

Selected metadata filter	Time to shortlist (sec)	Time to download (sec)	Total cell hits	Distinct cell groups	CSV File size (kB)
Species = Monkey AND Cell type = Principal cell	0.5	1	2,700	231	174
Brain region = Basal ganglia AND Cell type = Glia	0.5	0.5	984	18	19
Cell type = Microglia OR Oligodendrocytes	0.5	2.5	9,750	182	228
Developmental stage = Adult OR Young adult AND Cell type = Interneuron	0.5	4.0	16,911	3,126	3,606
Species = rat AND Brain region = Hippocampus AND Cell type = microglia	0.5	2	1,244	33	33
Cell type = GFAP-positive astrocytes AND	0.5	2	1,613	12	41

Slicing direction = coronal					
(Species = mouse OR rat) AND (Brain region = Neocortex OR hippocampus) AND Cell type = microglia	0.5	2.5	7,193	106	158
(Species = monkey OR mouse) AND (Cell type = astrocytes OR microglia) AND Label = immunostaining	0.5	2.5	4,296	144	131

To illustrate the content and utility of the CSV files, we employed summary reporting to extract the metadata grouping characteristics, morphometric values, and PDVs for all glial cells in the database as of version 8.0 (released 2020-06-29). This dataset includes 11,569 cells from 26 contributing laboratories, 8 animal species, 14 brain regions, and a variety of other experimental details, for a total of 218 groups (**Figure 3.4**). The format of the downloaded files makes it seamless to determine and quantify the diversity in the available data using standard spreadsheet programs such as Microsoft Excel. For example, from the metadata grouping CSV file (Fig. 3.4A, top) we derived the count and identities of differing metadata fields between all pairs of groups (Fig. 3.4A, bottom). Similarly, from the morphometric CSV file (Fig. 3.4B, top) we derived the count and identities of the morphometric parameters that displayed statistically significant pairwise differences between groups (Fig. 3.4B, bottom).

A.

	A	B	C	D	E	F	G	H	I	J	K
METADATA											
Fields	Title	1	2	3	4	5	6	7	8	9	
numofNeurons	24	26	28	26	28	27	70	15	15		
archive	Fulton	Fulton	Fulton	Kannan	Kannan	Kannan	Ohgomori	Ohgomori	Ohgomori		
species	mouse	mouse	mouse	rabbit	rabbit	rabbit	mouse	mouse	mouse		
strain	C57BL/6J x	C57BL/6J x	C57BL/6J x	New Zealand	New Zealand	New Zealand	C57BL/6J	C57BL/6J	C57BL/6J		
min_age	8	8	8	29	29	29	2	2.2	2.2		
max_age	11	11	11	29	29	29	5.2	2.2	2.2		
age_scale	Day	Day	Day	Day	Day	Day	Month	Month	Month		
min_weight	0	0	0	30	30	30	20	20	20		
max_weight	0	0	0	30	30	30	35	35	35		
note	Organotypic	Organotypic	Organotypic	aparotomy	aparotomy	aparotomy	None	None	None		
age_classification	neonatal	neonatal	neonatal	fetus	fetus	neonatal	adult	adult	adult		
gender	Not reported	Not reported	Not reported	Male/Female	Male/Female	Male/Female	Female	Female	Female		
brain_region	cerebellum	cerebellum	cerebellum	corpus callosum	corpus callosum	corpus callosum	spinal cord	spinal cord	spinal cord		
cell_type	Glia, Oligod	Glia, Oligod	Glia, Oligod	Glia, astrocy	Glia, astrocy	Glia, astrocy	Glia, microgl	Glia, microgl	Glia, microgl		
original_format	Neuronstuc	Neuronstuc	Neuronstuc	NeuroLucid	NeuroLucid	NeuroLucid	NeuroLucid	NeuroLucid	NeuroLucid		
protocol	culture	culture	culture	in vitro	in vitro	in vitro	in vitro	in vitro	in vitro		
slicing_thickness	350	350	350	30	30	30	50	50	50		
slicing_direction	parasagittal	parasagittal	parasagittal	coronal	coronal	coronal	horizontal	horizontal	horizontal		
stain	green	green	green	fluorescence	fluorescence	fluorescence	immunostain	immunostain	immunostain		
magnification	40	40	40	40	40	40	40	40	40	Not reported	Not reported

"1;2;3;4;6" indicate the different metadata fields between group 1 and group 6

Groups	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	1	22;	22;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;3;4;6;7;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
2	1	22;	22;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;3;4;6;7;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
3	1	1	22;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;3;4;6;7;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
4	6	6	7	22;	1;2;3;4;6;	1;2;3;4;6;	1;2;3;4;6;	1;3;4;6;7;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
5	9	7	5	5	4;6;7;8;9;	4;6;7;8;9;	4;6;7;8;9;	1;2;3;4;6;	1;2;3;4;5;	1;2;3;4;5;	1;2;3;4;5;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
6	5	11	9	7	10	22;	22;	1;2;3;4;6;	1;2;3;4;5;	1;2;3;4;5;	1;2;3;4;5;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;
7	6	6	7	16	12	1	22;	1;2;3;4;6;	1;2;3;4;5;	1;2;3;4;5;	1;2;3;4;5;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
8	8	6	9	18	10	7	9	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
9	16	7	20	19	9	10	15	12	22;	22;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
10	9	17	11	18	15	13	8	16	1	22;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
11	12	18	20	17	12	15	9	13	1	1	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
12	15	20	14	15	12	9	11	14	9	18	11	1;3;4;5;6;	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
13	16	7	20	19	9	10	15	12	12	8	16	5	1;3;4;5;6;	1;3;4;5;6;	1;2;3;4;5;6;
14	17	12	15	9	13	11	18	15	13	8	14	9	8	1	1

B.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
MORPHOMETRICS															
Group	Surface	Volume	Soma Surface	No. of Stems	Width	Height	Depth	Diameter	Euclidean Distance	Path Distance	Contraction	Length	Partition Asymmetry	Branch Order	
1	1108.6	287.82	68.574	4	46.19	85.94	4.86	0.56456	57.5818	77.7909	0.87357	570.51	0.664835	10	
1	2028.7	620.68	279.29	11	63.61	57	5.98	0.65507	43.4346	60.5145	0.908712	856.03	0.646276	14	
2	1768.4	512.45	113.53	10	42.47	76.33	3.8	0.6533	43.3628	66.333	0.906433	745.11	0.589515	12	
2	2125.6	467.63	315.18	17	56.14	72.33	5.49	0.60678	56.0214	91.7931	0.846336	1050.4	0.572988	14	
3	3003.9	786.38	163.98	4	45.32	78.87	4.99	0.68722	76.8537	90.7609	0.906946	1278.1	0.625853	21	
3	1608.8	461.46	84.075	9	37.53	49.94	5.34	0.73778	36.7378	45.8223	0.911546	629.13	0.484141	9	
3	1835.3	363.1	99.164	8	55.15	78.24	5.98	0.62732	54.3713	74.8027	0.890191	784.34	0.512646	9	
4	1895.2	374.16	295.9	19	31.38	48.25	5.02	0.61485	38.3313	48.0141	0.875648	655.18	0.602658	7	
4	2713.2	684.64	289.59	7	65.85	91.06	7.44	0.60872	71.3262	137.752	0.889927	1295.8	0.624083	12	
5	1003	360.58	204.99	14	27.48	31.09	3.31	0.63974	24.2465	32.3227	0.861435	399.64	0.556905	7	
5	3324.8	837.31	25.942	7	54.05	67.94	5.68	0.64538	59.9958	76.9053	0.868003	1481.5	0.550563	15	

"0" indicates the "Branch Order" is significantly different (Bonferroni-corrected t-test) between group 1 and 2

Groups	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	2	1	0	0	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
2	1	0	0	0	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
3	1	1	0	0	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
4	13	13	14	14	E:FO	E:FO	C:O	C:O	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
5	13	13	10	4	G:KL:N:O	E:FG:KL	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
6	13	13	15	3	7	0	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
7	13	13	15	2	8	3	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
8	13	13	8	10	10	11	12	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI	D:G:HI
9	13	13	12	11	15	12	11	12	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL	E:FG:KL
10	12	10	10	19	14	16	18	12	11	11	C:D:G:HI	C:D:G:HI	C:D:G:HI	C:D:G:HI	C:D:G:HI
11	9	7	5	16	16	17	17	9	15	12	C:D:G:HI	C:D:G:HI	C:D:G:HI	C:D:G:HI	C:D:G:HI
12	15	12	10	16	15	13	14	10	11	10	14	14	14	14	14
13	15	14	12	16	16	12	12	13	11	14	12	3	2	0	0
14	15	13	11	16	17	13	14	12	11	13	11	2	1	1	1

Figure 3. 4. Metadata and morphometry summaries. **A.** Different metadata groups (top) and corresponding pairwise comparison table (bottom). Colors indicate oligodendrocytes (yellow), astrocytes (green), microglia (blue), and other glia (pink). Red squares highlight an example pairwise comparison. In the comparison table, entries above the diagonal list the different metadata dimensions and entries below the diagonal indicate their count. **B.** Morphometric parameters for all cells organized by groups (top) and corresponding comparison table (bottom).

Based on the above data, we can then characterize the metadata and morphometric distinctiveness of each group of cells based on the extent by which that group differs from all other groups. Similarly, we can compute the distinguishing power of each metadata dimension and morphometric parameter as the proportion of pairwise group comparisons differentiated by that particular characteristic (**Figure 3.5**). On average, glial groups differ from each other by 18.9 ± 4.2 (mean \pm standard deviation) metadata features or $54.0\% \pm 11.9\%$ of the 35 possible annotated properties (Fig. 3.5A). The most distinctive subset of data in this regard is the control group of mouse glial cells from Denk archive (Helmstaedter et al., 2013b), which differs on average in 23 metadata details from other groups. The least distinctive subset is the control group of rat microglia from La Barbera archive (Krashia et al., 2019), which differs on average in 16 metadata details from other groups. Brain region is the most distinguishing metadata dimension, differentiating more than 96% of group pairs when considering all available anatomical information, such as sub-area and layer. Interestingly, glial cell types and subtypes (for instance, fibrous astrocytes vs. vessel-associated microglia vs. NG2-positive

oligodendrocytes vs. others) differentiate over three-quarters of the group comparisons (Fig. 3.5B). By contrast, the staining method has low distinguishing power (<30%), underscoring widespread usage of relatively uniform techniques (immunolabeling) across a broad diversity of biological analyses (nervous system location and cell phenotype).

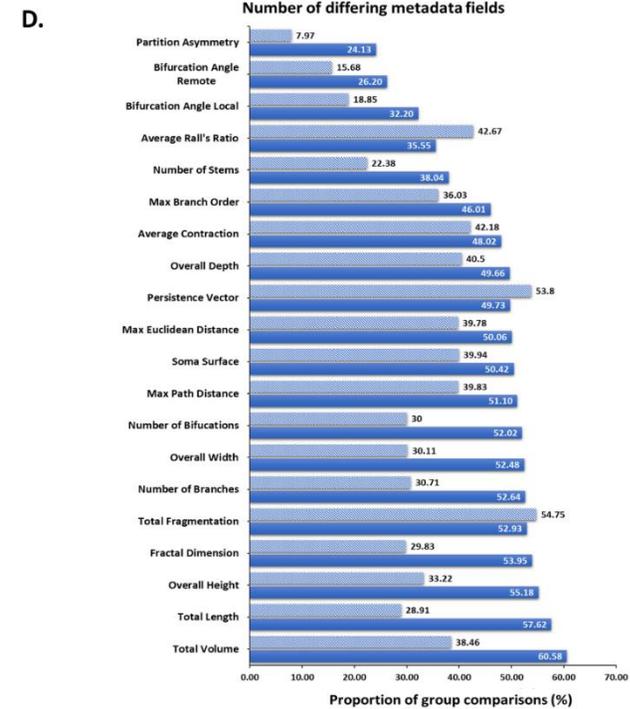
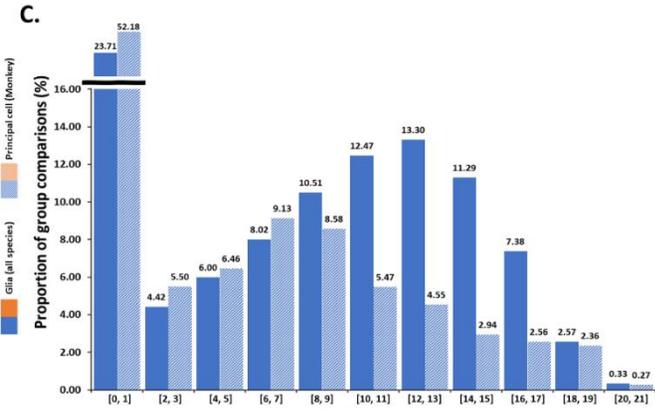
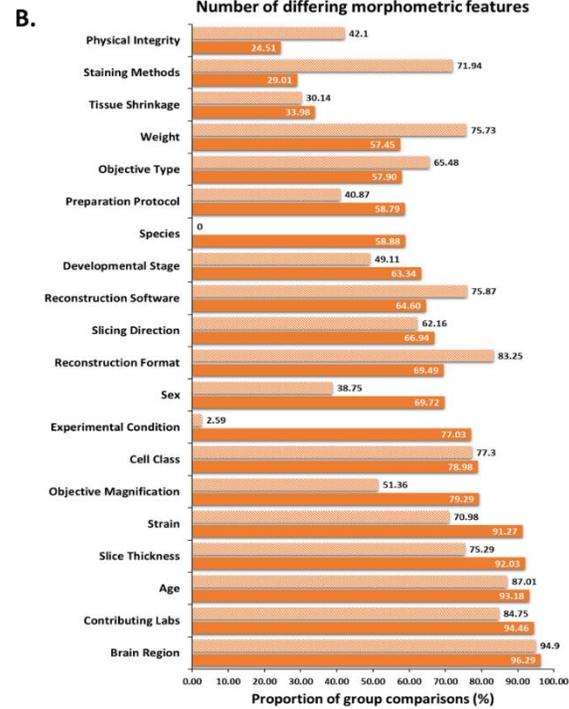
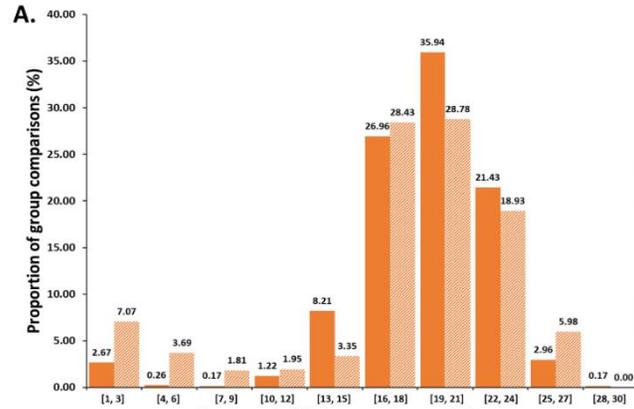


Figure 3. 5. Group distinctiveness and feature distinguishing power of NeuroMorpho.Org glia and monkey principal neurons. **A.** Histogram distribution of the average number of metadata details (out of 35 fields) differentiating a group from the other groups. **B.** Distinguishing power of the 35 metadata fields expressed as the percentage of group pairs differing in each dimension. **C.** Histogram distribution of the average number of morphometric parameters (out of 22) statistically differentiating a group from the other groups. **D.** Distinguishing power of the 22 morphometric parameters expressed as the percentage of group pairs differing in each parameter.

To investigate how representative these observations might be over different datasets, we repeated the analysis for all monkey principal neurons. These groups differ from each other by 17.5 ± 6.1 metadata features or $50\% \pm 17.4\%$ of the 35 possible annotated properties (Fig. 3.5A). The most distinctive subset of metadata is the control group of monkey pyramidal neurons from Wearne_Hof archive (Duan et al., 2002a), which differs on average in 23 metadata details from other groups. The least distinctive subset is the control group of monkey dLGN-projecting neurons from Briggs archive (Bragg et al., 2017; Briggs et al., 2016), which differs on average in 8 metadata details from other groups. Brain region is the most distinguishing metadata dimension followed by the age of the animals. As expected, species has a 0% distinguishing power as all the animals are monkey followed by the experimental condition (<3%) because almost all groups are control (Fig. 3.5B).

Just like for metadata, we next compared the 218 distinct glial groups with each other in terms of statistically significant differences in morphometric parameters. For this analysis, we considered the branch patterns captured by PDVs as an additional morphometric feature using arccosine statistical testing as described in the Methods section. On average, glial groups differ from each other by 8.7 ± 2.4 morphometric parameters (Fig. 3.5C) or $41.6\% \pm 11.5\%$ of the 22 possible measure variables (21 extracted by L-Measure plus PDVs). The most distinctive group is the lipopolysaccharide injection group of rat microglia from Balleine archive (Becchi et al., 2017), which differs on average in 13 morphometric parameters from other groups. The least distinctive group is the traumatic brain injury group of mouse microglia from Weil archive (Karelina et al., 2016), which differs on average in 3 morphometric parameters from other groups. We then calculated the distinguishing power of each morphometric parameter (Fig. 3.5D). The total volume occupied by glial processes has the highest distinguishing power and can differentiate over 60% of group pairs. In contrast, partition asymmetry (measuring the topological imbalance of arbors) and bifurcation angles are relatively uniform across glia, only differentiating between one-quarter and one-third of group comparisons. Interestingly, persistence diagram vectors display intermediate distinguishing power, being able to statistically separate almost half of the group pairs.

Lastly, the monkey principal neuron groups differ from each other by 4.26 ± 5.36 morphometric parameters (Fig. 3.5C) or $19.3\% \pm 24\%$ of the 22 possible measure variables. The most distinctive group is the control group from Fujita archive (Oga et al.,

2016), which differs on average in 11 morphometric parameters from other groups. The least distinctive group is the control group from Luebke archive (Amatrudo et al., 2012a; Rocher et al., 2010), which differs on average in 1 morphometric parameter from other groups. Total fragmentation had the highest distinguishing power among morphometric features by differentiating over 54.75% of group pairs followed by persistence diagram vectors with 53.8% distinguishing power. Partition asymmetry had the lowest distinguishing power of approximately 8% (Fig. 3.5D).

Discussion

Various functionalities have been continuously added since the launch of NeuroMorpho.Org 15 years ago to make the reuse of data efficient and user friendly, from the early morphological standardization pipeline (Halavi et al., 2008b), through the ontological search engine (Polavaram & Ascoli, 2017b), to the more recent implementations of extended file standards (Nanda et al., 2018a) and automated literature mining (Maraver et al., 2019a). Of particular relevance here, the Application Programming Interface allows machine-readable access to the database metadata and morphometric content. While the API can be used to download data in bulk programmatically, the absence of an intuitive graphical interface constituted until now an undeniable obstacle for non-technical users. To overcome this impediment, this work introduced summary reporting, a novel tool enabling end-users to seamlessly filter and download metadata and morphometrics information of interest through web-accessible dropdown menus.

Over the years, an entire ecosystem of distributed resources emerged in the computational neuroscience community that are either interoperable with or direct extensions of the data and functionality of NeuroMorpho.Org. For example, increasingly sophisticated modeling tools enable the investigation of single-neuron activity both by compartmental and stochastic-diffusion simulations (Edwards et al., 2014; Magalhães et al., 2019; Ray et al., 2016). Similarly, powerful approaches have been developed for the quantitative analysis of neural arbors, enabling rigorous statistical comparisons of datasets across species, brain regions, and experimental conditions (Ledderose et al., 2014; Lu et al., 2015; Torben-Nielsen, 2014). The growing recognition of the potential for novel scientific insight and discovery from secondary usage of digital reconstructions of neural morphologies further stimulated the creation of dedicated software utilities for efficient searching and mining of NeuroMorpho.Org data (Conjeti et al., 2016; Costa et al., 2016; Wan et al., 2015). Most recently, novel computer programs appeared that semi-automate the high-throughput query, retrieval, and testing of NeuroMorpho.Org tracings (Friedman, 2020; Mottini et al., 2015; O’Halloran, 2020). Despite this mounting abundance of applications, there is no similar tool available operating on neuromorphological files like summary reporting. The open-source addition of this new resource to the neuroinformatics workbench will help end-users identify the appropriate datasets to tackle the scientific questions of their calling.

With recent technological breakthroughs and expected continuous advancements in both microscopic imaging and 3D computer reconstructions, the accumulation of morphological data is noticeably accelerating. With the number and completeness of the available tracings, the diversity of the repository content increases as well: the recent introduction of glia in the database is only one of many examples of genuinely new classes of available data. Moreover, the switch to an agile release pipeline in 2020 implies the rapid-fire posting of often daily new datasets as soon as they clear the ingestion workflow after the publication of the corresponding research article. The free availability of such a large-scale resource undoubtedly benefits scientific progress. At the same time, its dynamic expansion also creates a challenge for researchers interested in meta-analysis, machine learning, and other big-science applications. The new NeuroMorpho.Org functionality described here, summary reporting, aims to solve this challenge by helping users handle the ever-rising complexity of data and metadata. With its intuitive ergonomics and user-friendly graphical interface, summary reporting complements and augments the existing NeuroMorpho.Org API for high-throughput data access to metadata, morphometrics, and persistence diagram vectors. Future work in the same direction includes the forthcoming roll-out of a new *similarity search* functionality enabling interested users to identify data of interest based on specific exemplars in the database.

Acknowledgments

The authors thank Andy Barthakur and Kunal Khurana for their assistance in data analysis, and Navy Merianda and Praveen Menon for software programming and debugging.

Funding

This work was supported in part by NIH grants R01NS039600, R01NS086082, and U01MH114829.

CHAPTER 4: MACHINE LEARNING CLASSIFICATION REVEALS ROBUST MORPHOMETRIC BIOMARKER OF GLIA AND NEURONAL ARBORS

Masood A. Akram, Qi Wei*, Giorgio A. Ascoli*

Center for Neural Informatics, Structures & Plasticity, Krasnow Institute for Advanced Study, and Department of Bioengineering, Volgenau School of Engineering, George Mason University, Fairfax, 22030, VA, USA.

* This chapter was posted on bioRxiv (2022), doi:
<https://doi.org/10.1101/2022.04.02.486839>

Acknowledgments. This work was supported in part by NIH grants R01NS39600, U01MH114829, and R01NS086082 to GAA, and R01EY029715 to QW. We are thankful to the NeuroMorpho.Org team and all data providers whose generous resource sharing made this study possible.

Conflict of Interest. The authors declare no conflict of interest.

Note: The supervised machine learning, discovery of ABEL, and manuscript writing is done by Masood A. Akram.

Abstract

Neurons and glia are the two main cell classes in the nervous systems of most animals. Although functionally distinct, neurons and glia are both characterized by multiple branching arbors stemming from the cell bodies. Glial processes are generally

known to form smaller trees than neuronal dendrites. However, the full extent of morphological differences between neurons and glia in multiple species and brain regions has not yet been characterized, nor is it known whether these cells can be reliably distinguished based on geometric features alone. Here, we show that multiple supervised learning algorithms (K-nearest neighbor, random forest, and support vector machine) deployed on a large database of morphological reconstructions can systematically classify neuronal and glial arbors with nearly perfect accuracy and precision. Moreover, we report multiple morphometric properties, both size-related and size-independent, that differ substantially between these cell types. In particular, we newly identify an individual morphometric measurement, Average Branch Euclidean Length (ABEL) that can robustly separate neurons from glia across multiple animal models, a broad diversity of experimental conditions, and anatomical areas, with the notable exception of the cerebellum. We discuss the practical utility and physiological interpretation of this discovery.

Keywords: Cellular Identity, Morphology, NeuroMorpho.Org, Neuroinformatics, Supervised Learning, Branch Length, Tree Size.

Introduction

Neuronal classification is an increasingly important subject because of its ultimate goal of linking cell types with computation, behavior, and cognition (Armañanzas & Ascoli, 2015c). The main experimental approaches to characterize neurons are

biochemistry, physiology, and morphology (Petilla Interneuron Nomenclature Group, 2008). These techniques have all yielded major breakthroughs in recent years thanks to rapid progress in genomics and transcriptomics, large-scale electric recordings, and high-resolution microscopic imaging (Litvina et al., 2019b), respectively. Both the European Human Brain Project and the American BRAIN Initiative identified cell type classification among their first priorities (Insel et al., 2013; Markram, 2012). Relative to neurons, glial cells have received less attention despite being similarly abundant in most organisms with a nervous system, including humans and all common animal models. Glia are involved in numerous important functions, such as myelination, anti-inflammatory protection, maintenance of neurochemical environment, and exchanges between nervous and vascular systems (Aguzzi et al., 2013b; Bronzuoli et al., 2018; Jessen, 2004; Rasband, 2016). Most glial cells emanate from the cell body complex branching processes that resemble the structural architecture of neuronal dendrites. While large numbers of neurons have been morphologically reconstructed for over three decades, digitally tracing glial trees has only more recently become a routine practice as well.

Although it is usually recognized that glial arbors are smaller than dendritic trees (García-Marín et al., 2007b; Y. Lu et al., 2015c; Veldman et al., 2020b; Zisis et al., 2021b), a comprehensive morphological comparison has not yet been carried out. In particular, it is still unknown whether these two main categories of cells can be reliably distinguished based on geometric features alone. The general problem is further complicated by several factors. First, both neurons and glia are intrinsically diverse, with

the former often distinguished by circuit role (long-range projecting, local interneurons, and sensory receptors) and the latter typically divided by functional specialization (microglia, astrocytes, oligodendrocytes, etc.). Second, both neurons and glia tend to differ broadly across animal species (especially between vertebrates and invertebrates), anatomical regions (e.g., neocortex, brainstem, spinal cord, peripheral nervous system), and developmental stage (such as embryo, early postnatal, and adult). Third, morphological characterization may be affected by the tremendous variability in experimental methods, including animal care, histological details, labeling protocol, imaging modality, and reconstruction software. Thus, it remains an open question whether suitable morphometric biomarkers exist that can robustly and systematically discriminate between neuronal and glial arbors.

Hundreds of laboratories worldwide continuously contribute their digital reconstructions of neurons and glia to the public online database NeuroMorpho.Org (Akram et al., 2018a). This repository associates every cell entry with metadata (Bijari, Akram, et al., 2020a) describing the animal subject (species, strain, sex, age, and weight), anatomy (brain region, sub-region, cell type, and sub-type), experimental details (protocol, condition, histology, microscopy, and tracing), and provenance (authors, source publication, original version, and processing logs). Moreover, the detailed 3D representation of arbor geometry is accompanied by a battery of morphometric parameters extracted with L-Measure (Scorcioni et al., 2008a), such as total length, number of branches, arbor height, and tortuosity. Glial cells were introduced to

NeuroMorpho.Org in version 7.1 (2017) and now constitute 11.3% of over 170,000 tracings. The unrestricted availability of these data provides an unprecedented opportunity for scientific exploration, statistical analysis, and computational modeling (Ascoli et al., 2017b).

Machine learning is a branch of artificial intelligence that aims at enabling the efficient and automatic detection of data patterns. Machine learning strives to produce the most accurate predictions, a distinct goal from that of statistical models designed to quantify the relationships variables (Aha et al., 1991b). Recent advancements in machine learning have not only benefitted healthcare with automatic diagnoses and treatment planning (Kohli & Arora, 2018), but were also successfully applied in neurobiological data analysis such as automatic tracing of neurons and glia (Peng et al., 2017c) and their quantification (Bijari et al., 2021b). In supervised machine learning, an algorithm is trained with the known class labels and identifies the most informative combination of features that are associated with those labels (Kotsiantis et al., 2006). The resultant classifiers can then be applied for predicting the labels of unknown data based on their feature values. Here we leverage supervised learning algorithms (K-Nearest Neighbor, Random Forest, and Support Vector Machine) to classify glia and neurons, and to recognize the morphological structures that distinguish these two main cell types of the nervous system.

Materials and Methods

Dataset selection and preprocessing. Morphological reconstructions of neurons and glia were obtained from NeuroMorpho.Org using the *Summary Reporting* web-based functionality (Akram et al., 2022b). This tool collates for every digital tracings the

annotation of 35 distinct metadata fields, providing a detailed qualitative description of the cell (Parekh et al., 2015c), as well as 21 morphometric measurements which capture the quantitative structural features of the arbor (Scorcioni et al., 2008a). First, we downloaded all glial cells available at the time we began data analysis (Fall 2020), corresponding to 10 consecutive releases of NeuroMorpho.Org (versions 7.1 to 8.0 inclusive). We then analyzed the distributions of their metadata with respect to animal species, developmental stage, anatomical region, and other experimental conditions, and queried the database to identify the same number of neurons with the most similar metadata characteristics. Since we were interested in comparing glial processes specifically to neuronal dendrites (as opposed to axons), only neurons with dendritic tracings available were selected. Moreover, we solely included neurons and glia with complete or moderately complete reconstructions, thus excluding those annotated as incomplete dendrites or incomplete glial processes by the original contributors. The resulting balanced dataset of 22,792 cells was comprised of 11,398 neurons and 11,394 glia.

Of the 21 morphometrics extracted for each cell from NeuroMorpho.Org, we excluded Soma Surface and Depth from the analysis. Soma Surface is not an arbor morphometric, and 4.7% of the tracings in our dataset did not include soma reconstruction. Depth was similarly not reported for 8.6% of the cells as the accuracy of the tracing is reduced in certain cases by light diffraction and tissue shrinkage in the direction perpendicular to the imaging plane. The remaining 19 morphometric features

were used in the analysis: number of stems, number of bifurcations, number of branches, overall width, overall height, average diameter, total length, total surface, total volume, maximum Euclidean distance, maximum path distance, maximum branch order, average contraction, total fragmentation, partition asymmetry, average Rall's ratio, average bifurcation angle local, average bifurcation angle remote, and fractal dimension. The formal definitions of these metrics are available on the Frequently Asked Questions of NeuroMorpho.Org (<http://neuromorpho.org/myfaq.jsp>) and on the online manual of L-Measure (<http://cng.gmu.edu:8080/Lm/help/index.htm>).

Most reconstructions in NeuroMorpho.Org have coordinates reported in microns. In a subset of reconstructions, however, the coordinates are expressed in pixels. In these cases, the nominal measurements listed in the morphometric tables must be converted by an appropriate scaling factor. Therefore, we manually calculated the height of at least one cell in each archive from the figures of the corresponding original publications (and relative scale bar) and compared the resulting value to the height reported by NeuroMorpho.Org. If the values did not match, we computed a conversion factor and applied it to size-related morphometric features including width, height, total length, total surface, total volume, maximum Euclidean distance, and maximum path distance. The specific archives that underwent rescaling and the calculations for the scale correction are detailed in the Supplementary Material at https://github.com/Masood-Akram/Classification_Neurons-Glia/tree/main/Supplementary_Material

When we concluded the main analysis for this work (Fall 2021), a new version of NeuroMorpho.Org had been released (v.8.1). We thus identified an additional balanced dataset of 4292 neurons and 4286 glia, up to version 8.1.90 (December 2021), allowing us to test the robustness of our main results on a completely independent dataset. The scale correction details and the complete metadata breakdown for this additional dataset are also included in the Supplementary Materials at https://github.com/Masood-Akram/Classification_Neurons-Glia/tree/main/Supplementary_Material

Dimensionality reduction. We computed the coefficient of determination (R^2) to quantify the pairwise correlation (Di Buccianico, 2008) among the 19 morphometric parameters across neurons and glia using the *rcorr* function in the R package *Hmisc* (Harrel & Dupont, 2021). We then used Principal Component Analysis (PCA) to reduce the feature redundancy. PCA transforms the data into a set of new orthogonal variables by identifying the directions (principal components) along which the variation in the data is maximal (Abdi & Williams, 2010). By discarding the least informative components, each sample can be represented by fewer, linearly independent features instead of more, mutually dependent variables (Ringnér, 2008). Thus, PCA reduces the dimensionality while retaining most of the variation of the data. PCA was performed with the R package *stats* (Core Team, 2021) by using the function *prcomp()* and setting *scale = TRUE*. Along with PCA, all parameters were standardized by first subtracting the mean of the entire feature vector from each element and then dividing by the standard deviation.

Supervised learning. Training data consisted of the normalized principal components of the morphometric features as input, and the known class labels (the cell identity ‘neuron’ or ‘glia’) as output. We used three distinct classification algorithms implemented in the R programming language (Core Team, 2021) v.4.1.1 for Windows. In all cases we calculated sensitivity, specificity, and accuracy, respectively defined as the fractions of true positives, true negatives, and correctly classified (true positives plus true negatives) cells, using the *caret* (Kuhn, 2021) package in R (Core Team, 2021).

K-Nearest Neighbor (KNN) is a supervised learning algorithm that can be used both for classification (discrete value output), as applied here, and regression (continuous value output) problems. In KNN, the training instances are stored with their labels and each new instance is compared with the labeled ones using a similarity matrix. The vote for each new instance’s label by comparing to existing instances is taken from the value of k . For example, if k is set to 5, 5 nearest neighbors are identified from the training instances and the class label with the highest frequency is assigned to the new instance (Aha et al., 1991b). The default Euclidean distance was used here to compute similarity between two data points. The built-in *caret* package (Kuhn, 2021) was utilized for the KNN implementation by using the *train()* function and setting *method* = “*knn*” with *tuneLength* = 10 and $k = 5$.

Support Vector Machine (SVM) is a binary classification algorithm based on finding the maximum margin hyperplane that gives the greatest separation between the data

points of different classes in multidimensional space. Those data points closest to the hyperplane are called the support vectors. If the data are not linearly separable, different kernels can be selected for nonlinear classification. This classifier is robust to large number of variables and small sample sizes (Cortes & Vapnik, 1995b). We implemented SVM using the *caret* package (Kuhn, 2021) using the *train()* function with *tuneLength = 10*, and *method = "radial"* kernel, which gave the best classification accuracy and is also a common choice for classification tasks (Luts et al., 2010).

Random Forest (RF) consists of a large number of individual decision trees. Each individual tree in the forest splits out a class prediction and the most frequent class becomes the model prediction. This is one of the most popular machine learning algorithms and is capable of both classification (as used here) and regression (Breiman, 2001b; Sarica et al., 2017a). We applied the *randomForest* package (Liaw & Wiener, 2002) using function *train()*, *method = "rf"*, and *ntree = 500*. The rationale for this choice is that a relatively high number of decision trees ensures that every input row is predicted multiple times. Parameter *mtry* determines the number of variables randomly sampled as candidates at each split and was set to the default value of 5.

K-fold Cross Validation (K-fold CV). It is customary in supervised learning to train the model on the majority of the data, leaving the remaining for testing. To rigorously examine the classification performance on our data, we performed K-fold cross validation. This process divides the dataset into k equal parts. A classifier is first trained

on $k-1$ parts for each fold. The accuracy of the trained model is then assessed by using the part of data excluded from the $k-1$ parts in training (Bouckaert, 2003). We performed 10-fold CV repeated 10 times using the *caret* (Kuhn, 2021) package by using the function *trainControl()*, *method = "repeatedcv"*, *number = 10*, and *repeats = 10*.

PSwarm is a global optimization solver for bound and linearly constrained problems (Vaz & Vicente, 2009). This algorithm is based on a pattern search and particle swarm method, which guarantees the convergence to stationary points from arbitrary starting points. We used the PSwarm Solver (v.1.5, June 2020, norg.uminho.pt/aivaz/pswarm/) implementation in R to find the linear discriminant of neurons and glia based on two morphometric parameters. We set the lower and upper bounds to 75 and 175, respectively, for intercept and to -10 and 75 for slope, and the number of iterations (*maxit*) to $2 \cdot 10^9$.

All analyses were carried out on a 64-bit machine equipped with an Intel Core i7-8565U and 16 GB of RAM running Windows 10. The R scripts utilized in this work are released open source at https://github.com/Masood-Akram/Classification_Neurons-Glia/tree/main/R_Code.

Average Branch Euclidean Length (ABEL) is the average over all branches in a cell of the straight-line distance between the beginning and ending points of each branch. This quantity was calculated from three of the morphometric parameters provided for each cell

by NeuroMorpho.Org: branch path (geodesic) length, the number of branches, and contraction, which is the ratio between Euclidean branch length and branch path length (its inverse is tortuosity). Specifically, ABEL was derived by summing the product of contraction by branch path (geodesic) length and then dividing the result by the total number of branches in each cell:

Equation 4. 1

$$ABEL = \frac{\sum_{i=1}^{NB} (Contraction \times Branch Path Length)}{NB},$$

where NB is the total number of branches. We also calculated ABEL of the terminal branches (from a bifurcating point to the tip) and of internal branches (between two consecutive bifurcation points) of both glia and neurons from the .swc reconstruction files provided by NeuroMorpho.Org using L-Measure. In particular, for every cell we first extracted path length and contraction values for each branch while setting “Terminal_Degree=1” under Specificity for terminal branches and “Terminal_Degree>1” for internal branches. We then multiplied the path length and contraction values and took the average within each group (terminal, internal). Lastly, we were also interested to determine the classification power of ABEL when only a small sample of branches was used to estimate the ABEL value. To this aim, we first extracted for every cell path length and contraction values of all branches with L-Measure without setting any Specificity (thus including both internal and terminal branches) and multiplied each pair of values to obtain the Euclidean lengths of all branches. We then utilized the *random* library (Van Rossum, 2020) in Python 3 (Van Rossum & Fred L., 2009) to stochastically select 100 sets of N values without replacement, where N varied from 1 to 15. The N values were

used to compute ABEL within each set, and the average and standard deviation were then computed over the 100 sets. Finally, classification was carried out using the mean ABEL value. The code for this analysis is released open source at: https://github.com/Masood-Akram/Classification_Neurons-Glia/tree/main/Python_Code.

Average Branch Euclidean Length (ABEL). All morphological reconstructions utilized in this work are available at NeuroMorpho.Org. The individual archive listing, the metadata, and the scaling adjustments are all included in the Supplementary Materials. The source code for all analysis is publicly available on github.com

Results

The morphological reconstructions of glial processes and neuronal dendrites utilized in this work were contributed to NeuroMorpho.Org by over 250 independent laboratories (listed in Supplementary Materials) and reflect the distribution of published arbor tracings in neuroscience

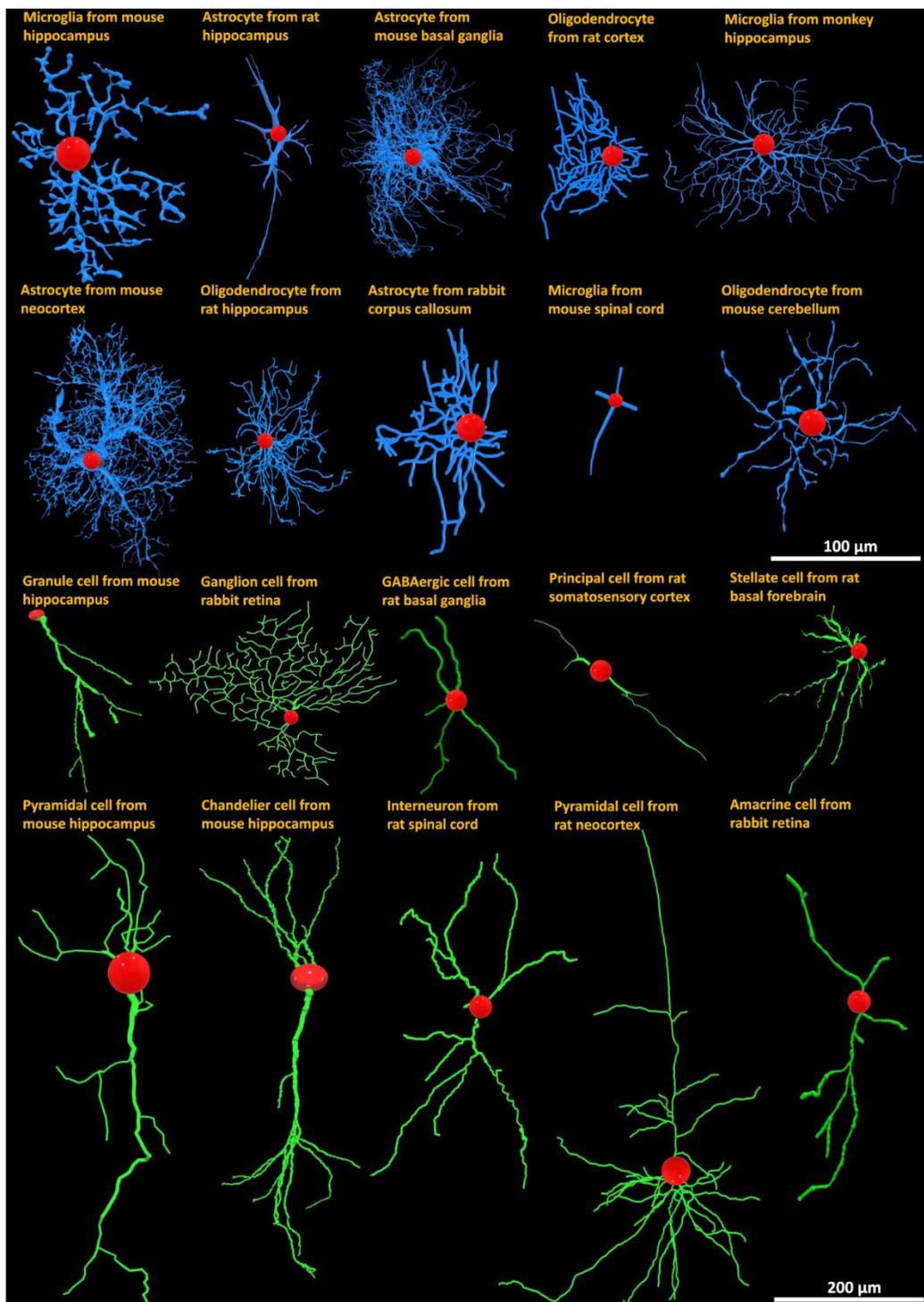


Figure 4. 1. Representative diversity of morphological reconstructions of glia and neurons from NeuroMorpho.Org with labels indicating animal species, anatomical region, and cell type. Blue: glial processes; green: neuronal dendrites; red: cell bodies.

(Fig. 1). Consistent with this multifarious provenance, the dataset spans a broad diversity of experimental methodologies, including over 20 different staining methods (e.g., genetic green fluorescent protein labeling, intracellular biocytin injection, immunostaining, and rapid Golgi), 15 digital reconstruction software (Neurolucida, Imaris, Amira, NeuronJ, Simple Neurite Tracer, Vaa3D, Knossos, NeuTube, etc.), and a continuum of ages across the developmental, from embryo through juvenile to old adults. Moreover, the data came from both mammalian and non-mammalian species and a large variety of anatomical regions but were balanced between neurons and glia across these dimensions (Fig. 2). The full breakdown of all metadata categories annotated in NeuroMorpho.Org is provided in Supplementary Materials.

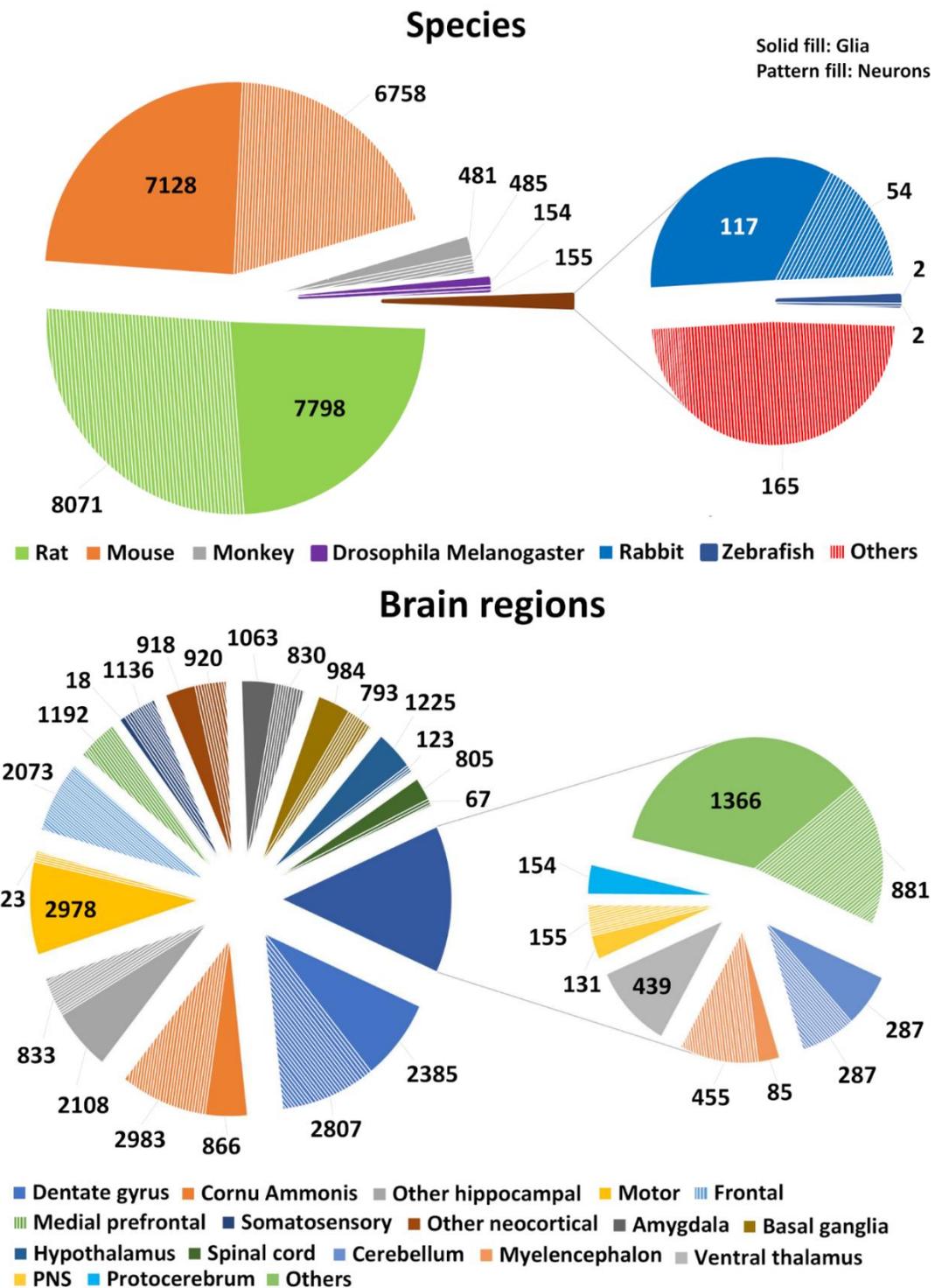


Figure 4. 2. Balanced distribution (A) animal species and (B) brain regions for the analyzed glia and neuron datasets.

The morphometric quantification of neural trees supplied by NeuroMorpho.Org provides a detailed 3D representation of branch geometry (Fig. 3). The extracted features include parameters characterizing both the overall size of the arborization and the scale-invariant properties. The formers include total cable length and surface area, spanning height and width, maximum Euclidean and path (geodesic) distance from the root (soma), and average branch diameter among others. The latter measurements capture bush complexity (e.g., number of branches and tree stems), branch angles (local and remote bifurcation amplitude), topological imbalance (partition asymmetry and maximum branch order), and spatial meandering (contraction and fractal dimension), among others. Altogether, this set of morphometric parameters is well suited to characterize the structure of neuronal dendrites and glial processes alike, and thus to quantify their similarities and differences.

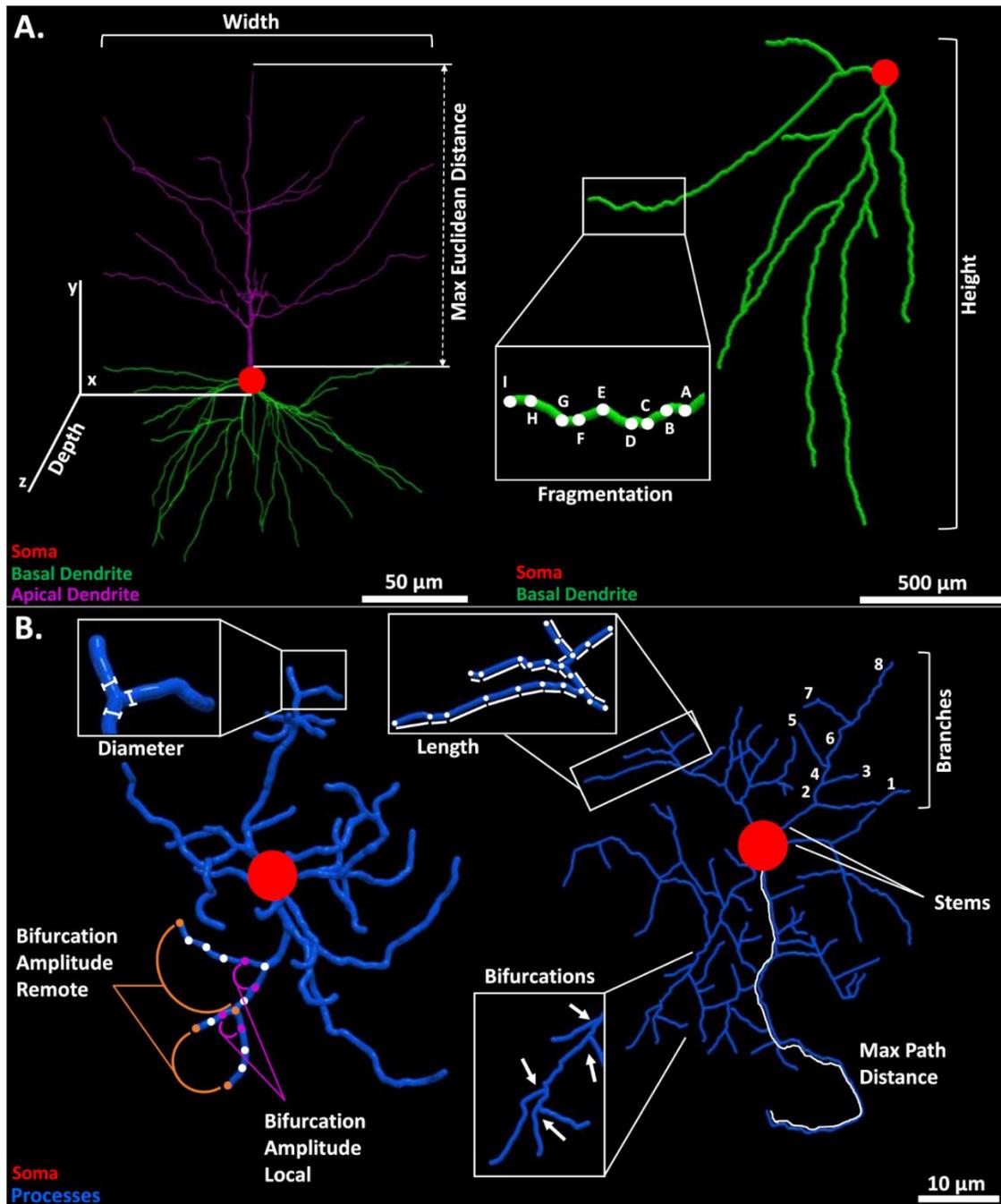


Figure 4.3. Schematic of selected morphometric features. (A) Illustration of width, depth, and maximum Euclidean distance (left) in a monkey neocortical pyramidal cell (NMO_00002) from the Wearne_Hof archive (Duan et al., 2002); and of height and fragmentation (right) in a hippocampal granule cell (NMO_73103) from the Diaz archive (Sebastián-Serrano et al., 2016). (B) Diameter and local or remote bifurcation amplitude (left) in a rat neocortical microglia (NMO_95641) from the Roysam archive (Megjhani et al., 2015b); and maximum path distance, length, and number of branches, bifurcations, and

stems in a rat cortical oligodendrocyte (NMO_131081) from the Sato_Bigbee archive (Mohamed et al., 2020).

Although the above-described parameters are intuitively interpretable, they may not be completely independent of each other. For example, total tree length, surface area, and average branch diameter are expected to be interrelated. This information redundancy can unduly bias the objective characterization of the structural differences between glia and neurons, complicating subsequent interpretations. The pairwise coefficients of determination (R^2) for glial (Fig. 4A) and neuronal (Fig. 4B) morphometrics confirm the substantial correlation between specific features. For example, surface is highly correlated to volume, the number of bifurcations to the number of branches, length, fragmentation, and branch order (and the latter four to one another), path distance to Euclidean distance, and contraction to fractal dimension. Although the coefficients of determination tended to be higher in neurons than in glia, most visibly between maximum path distance and total surface area, and between overall height and maximum Euclidean distance, the majority of correlations were highly consistent between the two cell types. In order to remove the interdependency among features, we performed PCA jointly on the full dataset to orthogonalize the morphometric parameters (Fig. 4C). The first 11 principal components captured 95.70% of the variance and we thus decided to exclude the last 8 components from machine learning. The 11 principal components considered in subsequent analysis constitute a combined transformation of all 19 morphometric parameters described above but are guaranteed by PCA to be independent.

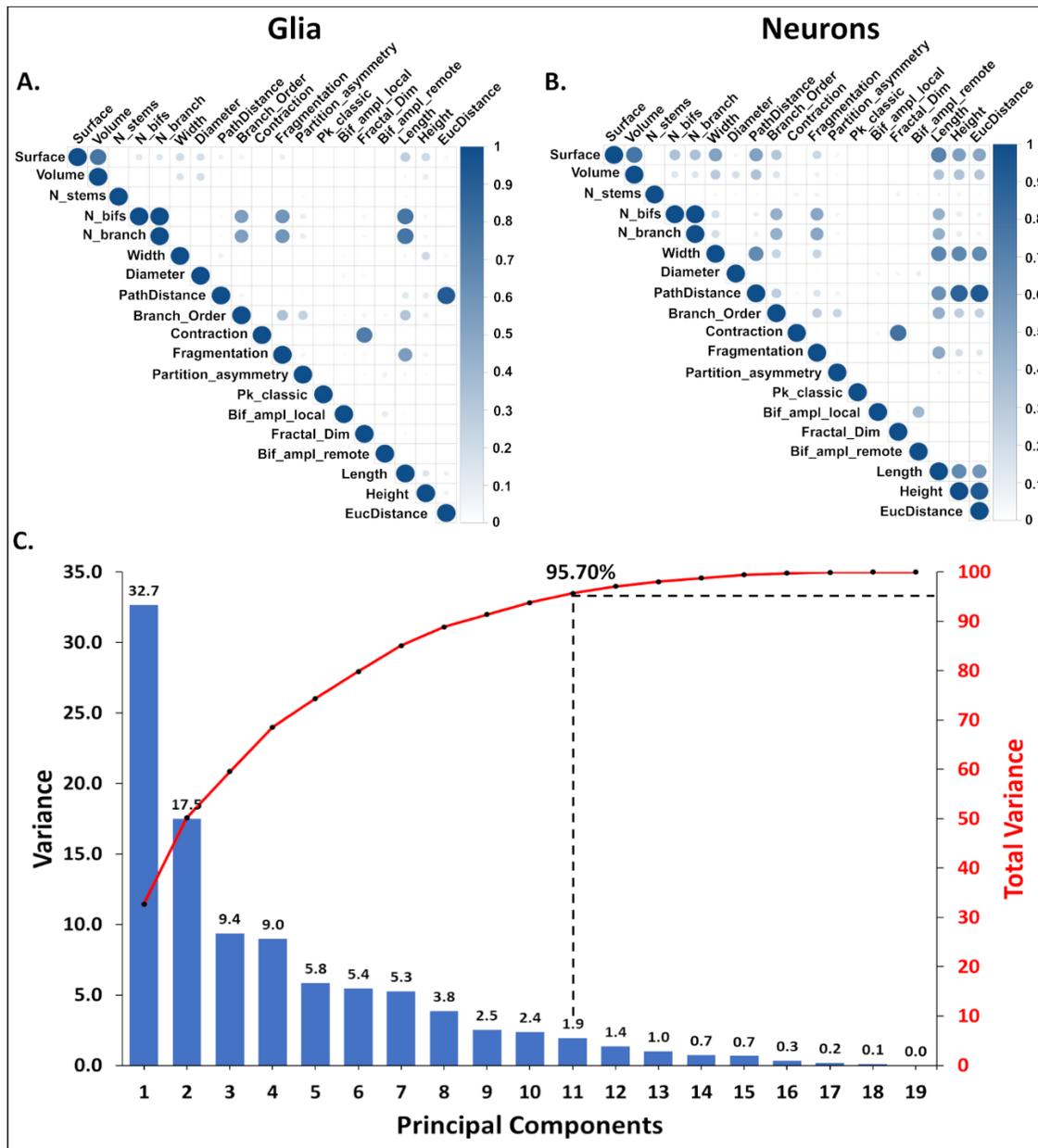


Figure 4.4. Orthogonalization of morphometric features. (A) Correlation matrix quantifying the interdependence among 19 morphometric features of glia and of (B) neurons. The coefficient of determination (R^2) is shown on a dark intensity scale. (C) Screen plot of the variance contributed by each sequential principal component (blue bars, left axis) and the corresponding cumulative distribution (red line, right axis).

The first two components (PC1 and PC2) alone capture more than 50% of the overall morphological variance in neural cells. A striking separation between cell classes is apparent on the PC1-PC2 projection plane, with neurons more abundant towards positive coordinates and glia towards negative in both dimensions (Fig 5A). Data points that are close to each other in this projected space represent structurally similar cells, whereas morphologically different cells occupy distant positions. The first two principal components consist of distinct linear combinations of morphometrics: PC1 (Fig. 5B) has strongly positive loading on size (e.g., total cable length, overall arbor height, maximum path distance), while PC2 (Fig. 5C) has strongly negative loadings on tree complexity and other scale-invariant measures, such as number of bifurcations, maximum branch order, and fractal dimension. These results therefore confirm that neurons have greater overall arbor size than glia, as quantifiable by multiple alternative metrics. Furthermore, this analysis reveals that, compared to neuronal dendrites, glial processes tend to form bushier trees, with more symmetric branching distributions and wider bifurcations angles.

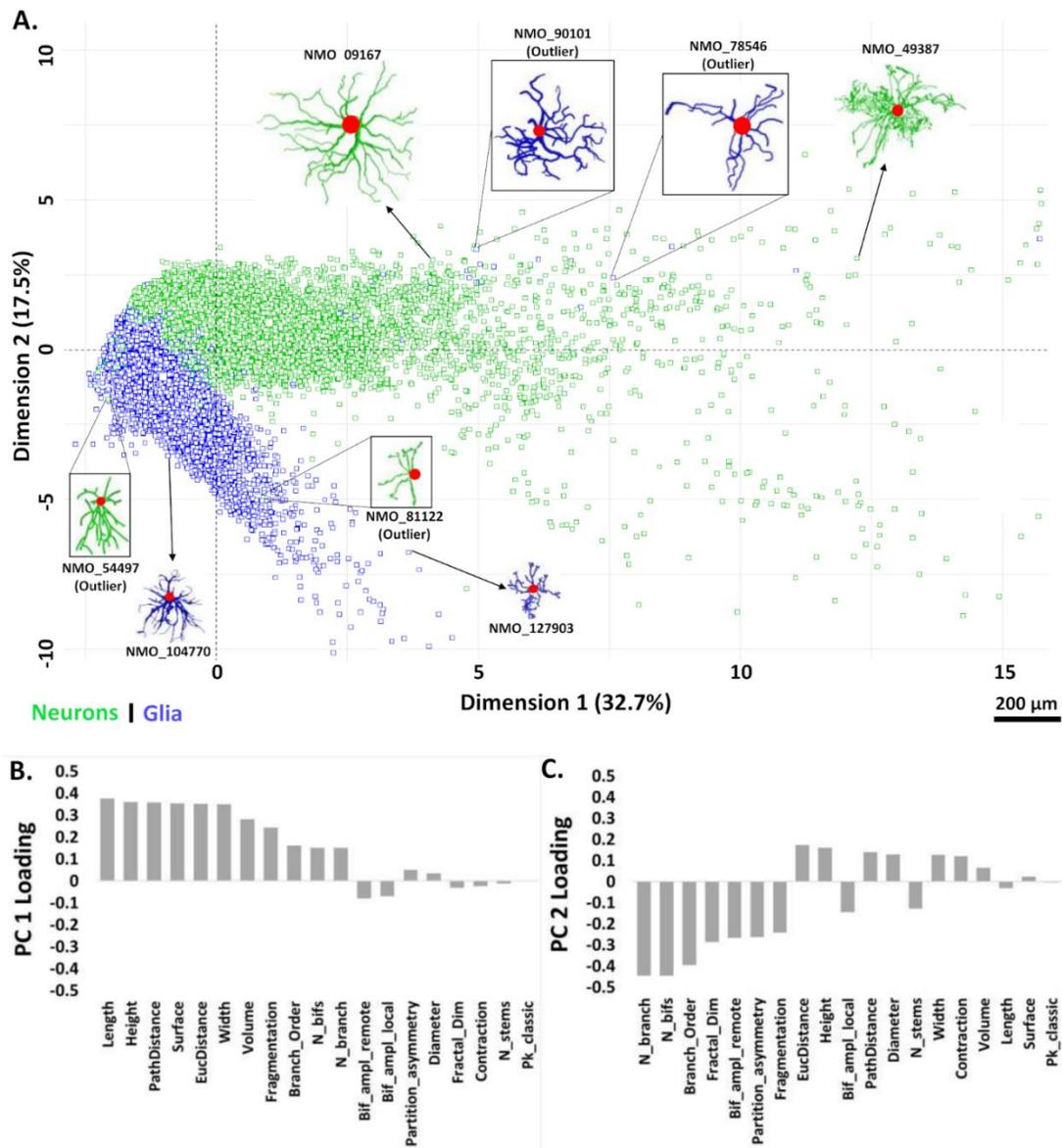


Figure 4. 5. (A) PCA biplot of the 2-dimensional distribution of neurons and glia relative to the first two principal components (PC1 and PC2). Morphological tracings of several cells (glia: blue; neurons: green) are also shown to illustrate their structural variability and similarity in this space. (B) Linear contributions of all morphometric parameters to PC1 and (C) PC2. Negative loadings indicate a high weight of the scale low-end for a parameter: for instance, cells with large positive PC2 values tend to have very few branches, whereas cells with many branches tend to have large negative PC2 values.

The above analysis suggests that neurons and glia may be reliably recognized based on morphological features alone independent of numerous confounds such as species, anatomical region, and experimental methods. To test this hypothesis, we used the 11 principal components explaining >95% of the variance for classification with three supervised learning algorithms: Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF). In all cases we performed 10-fold cross validation: the dataset was randomly split into 10 folds without replacement, with 90% of the data used to train the classifier and the remaining 10% used for testing. The process was repeated 10 times for more reliable assessment. The total runtime for 10 repeats of 10-fold cross validation was 15 minutes for KNN with 5 nearest neighbors ($k=5$), 2.2 hours for SVM, and 5 hours for RF. All three classifiers performed remarkably well in separating glia from neurons (Fig. 6). SVM slightly outperformed KNN in terms of sensitivity, specificity, and accuracy, with RF displaying intermediate performance metrics. However, all classification measurements fell within 1% difference for the three algorithms, and the area under curve (AUC), a robust measure of predictive modeling accuracy, was >99.5% for each of them.

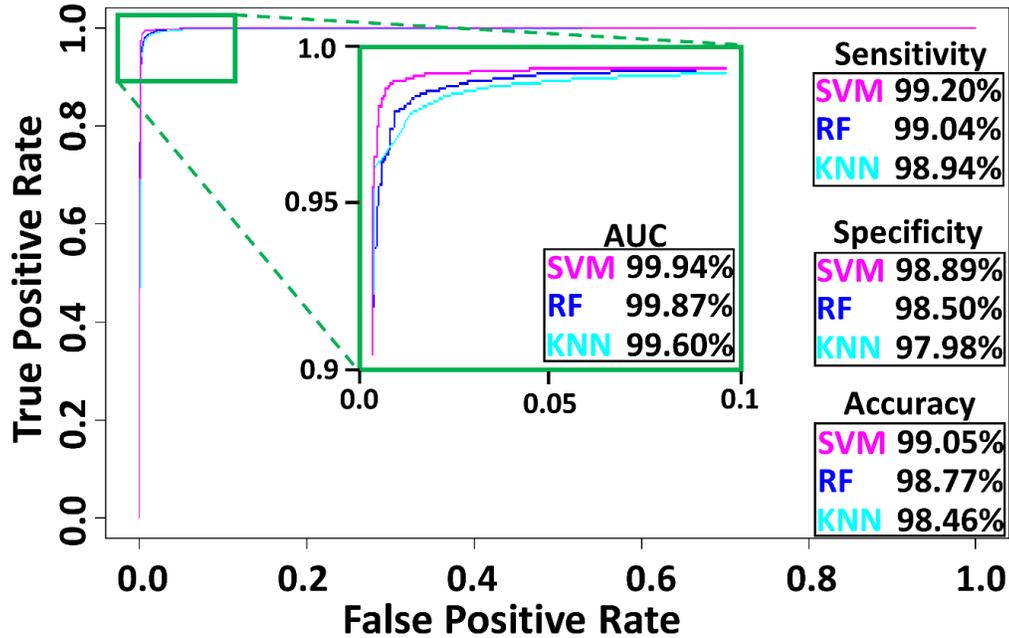


Figure 4. 6. Classification performance for Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF), including the area under the curve (AUC) of the Receiver Operating Characteristic plot.

The supervised classification results clearly demonstrate that a variety of automated methods can reliably distinguish glia from neurons by using morphological features alone. Given that the principal components utilized to train the machine learning algorithms represent an extensive battery of morphometric measurements, the question remains whether individual geometric features can be identified that achieve similarly robust performance. Based on the PC1 and PC2 loadings described above, measures of arbor size such as overall height and total length and measures of arbor complexity such as number of branches could constitute viable candidates. Although the corresponding silhouette profiles (Fig. 7) corroborated the expected statistical differences, it also

revealed extensive overlap in the corresponding data distributions. For example, the optimal height threshold to discriminate neurons from glia (76.15 μm) resulted in a suboptimal classification accuracy of <0.95 , with $>6\%$ of neurons misclassified, and even worse performance for total length, number of branches, and any other individual parameter. We reasoned, however, that since neurons have longer cable and glia have more branches, an appropriately combined feature could achieve a multiplicative improvement in discrimination. Specifically, dividing length by number of branches, which defines average branch path length, should yield parameter values with an even larger ratio between neurons and glia than length alone. Moreover, since neurons have slightly less tortuous branches than glia, as indicated by larger (if only marginally) contraction values, multiplying branch length by contraction and averaging over all branches, which defines average branch Euclidean length (ABEL), should further increase corresponding parameter value between neurons and glia. Silhouette analysis confirms the considerably better separation between neurons and glia based on ABEL when compared to all other individual morphometrics (Fig. 7).

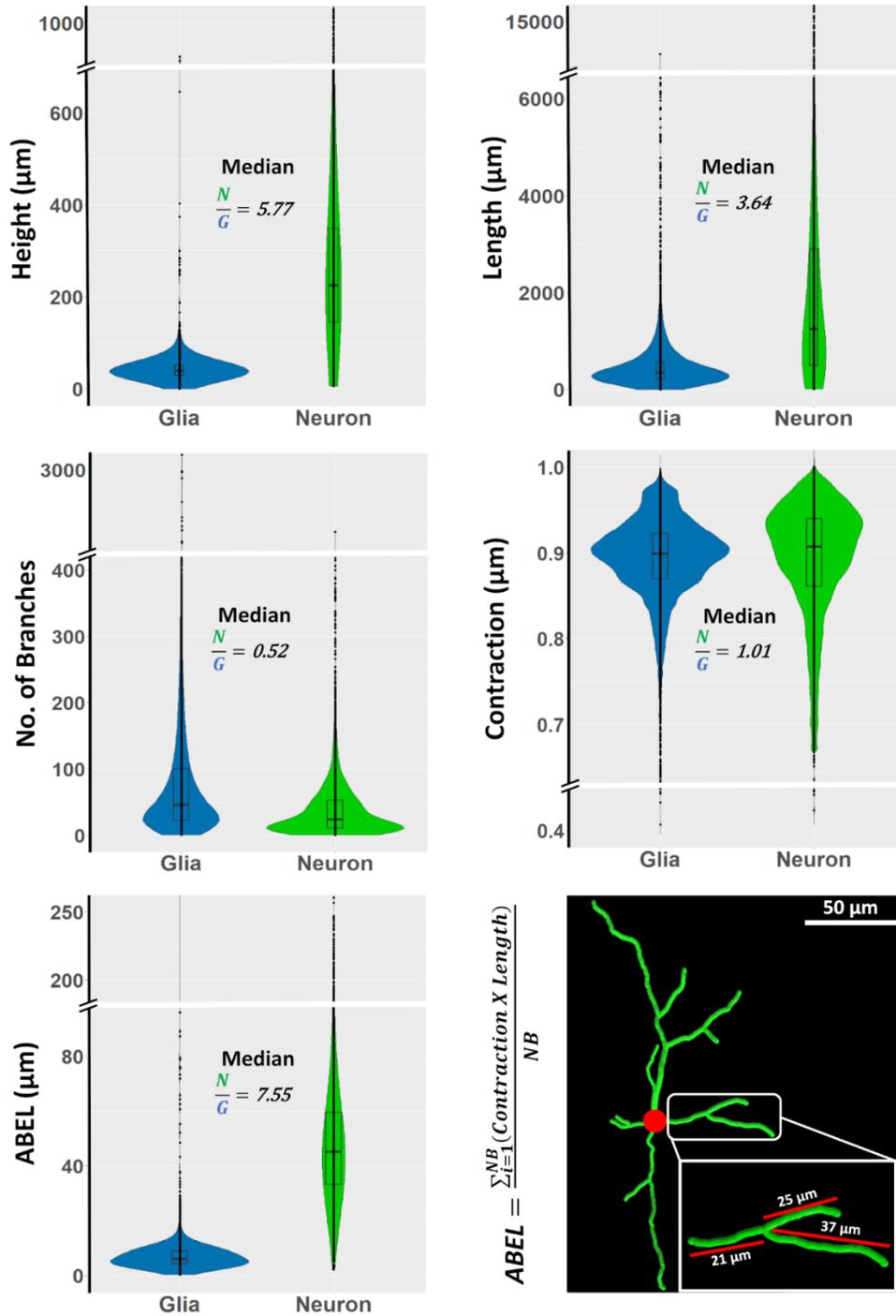


Figure 4. 7. Silhouette profiles of length, height, contraction, number of branches, and average branch Euclidean length (ABEL) of glia and neurons, and examples of branch Euclidean length measurements from a rat basal ganglia GABAergic cell (NMO_68194) from the Smith archive (Smith et al., 2015).

The optimal ABEL threshold of 14.33 μm results in overall classification accuracy of 97.6%, with fewer than 2.4% of glia and 2.5% of neurons misclassified. Notably, the misclassification rate dropped steeply around the threshold ABEL value, with >90% of the misclassified cells found in a narrow ABEL range of 7 (12-19) μm (Fig. 8A). These results were robust across multiple species, strains, developmental stages, anatomical regions, types of glial and neuronal cells, labeling techniques, and experimental methods, as detailed in the Supplementary Materials. For example, when dividing all data by contributing labs, for more than three-quarters of cases the misclassification rate was less than 1%. The rare exceptions consisted of specific phenotypes as discussed at the end of the Results. Furthermore, even an incomplete sampling of neural branches is sufficient for reliable classification based on ABEL: the accuracy is essentially unaltered when using 15 randomly chosen branches (97.3%) and remains above 95% when reducing the ABEL sample size to 5 branches (Fig. 8B). To determine if the classification could be improved further by considering arbor height together with ABEL, we combined the two measures (Fig. 8C). The optimal linear boundary separating neurons and glia followed the equation $A = -0.1352H + 23.04 \mu\text{m}$, where A and H stand for ABEL and height, respectively. This combination increased the classification accuracy of glia and neurons only marginally compared to using ABEL alone, from 97.6% to 98.5%. Altogether, these results indicate that ABEL is an effective, *novel morphological biomarker* for identifying the main neural cell class.

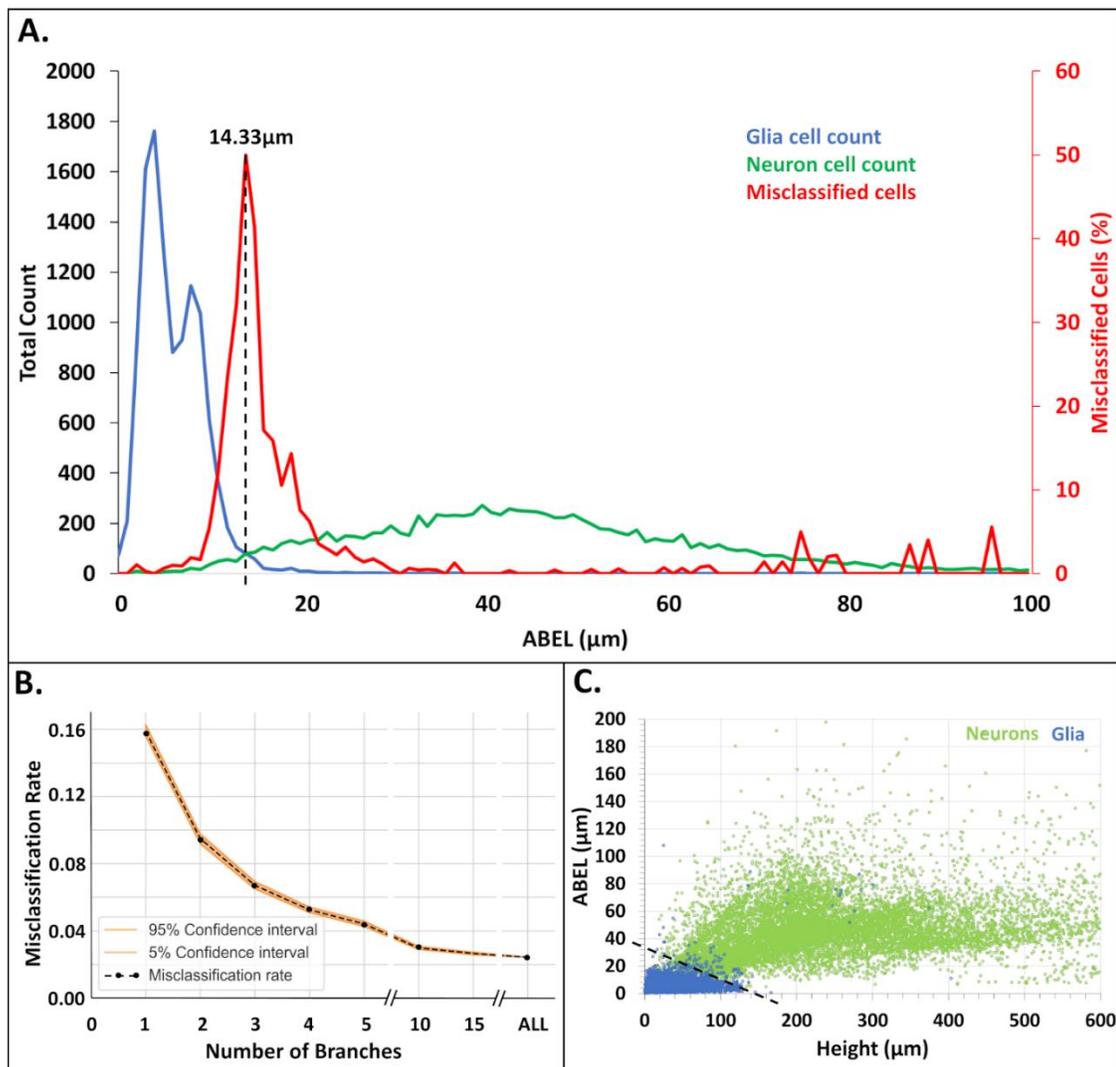


Figure 4. 8. Classification performance of average branch Euclidean length (ABEL). (A) ABEL distributions of neurons (green), glia (blue), and cells that are misclassified (red, secondary axis) based on optimal separation threshold of $14.33\ \mu\text{m}$ (vertical dashed line). (B) Misclassification rate as a function of the number of branches sampled to estimate ABEL. (C) Linear separation (black dashed line) between neurons (green) and glia (blue) on the plane defined by arbor height and ABEL.

Multiple studies reported that certain neuron types have longer terminal branches than internal (bifurcating) branches (Duan et al., 2002b; Kawaguchi et al., 2006; Y. Li et al., 2005), but it is unknown whether the same may be true for glia. Since neurons have greater ABEL values than glia, if glial processes have similar length for their terminal and internal processes, then terminal ABEL might be even more effective than overall ABEL to distinguish neurons from glia. To test this possibility, we extracted terminal and internal ABEL for all cells. The distribution of the ratios between terminal and internal ABEL values had an average of approximately 2 for neurons (Fig. 9A), confirming earlier reports that dendrites tend to have longer terminal than internal branches. In contrast, the distribution of the terminal/internal ABEL ratios had an average close to unity for glia, indicating that this phenomenon is limited to neurons. This was also confirmed by linear regression analysis, where the relationship between terminal ABEL and overall ABEL was essentially described by the identity line for glia, but had a slope above unity for neurons (Fig. 9B). Nevertheless, terminal ABEL did not improve the classification accuracy of glia and neurons compared to overall ABEL: in fact, it was slightly decreased to 97.1%, with an optimal separation threshold of 16.20 μm . To investigate why restricting ABEL measurements to terminal branches failed to improve classification performance, we examined the terminal/internal ABEL ratio specifically for the misclassified cells (Fig. 9C). Interestingly, those outlying neurons with exceptionally low ABEL values also displayed similar length between terminal and internal branches. Conversely, outlying glia with exceptionally high ABEL values had longer terminal than internal branches. Linear regression of terminal ABEL versus overall ABEL for the

misclassified cells also confirmed that most cells misclassified using overall ABEL are also misclassified using terminal ABEL (Fig. 9D).

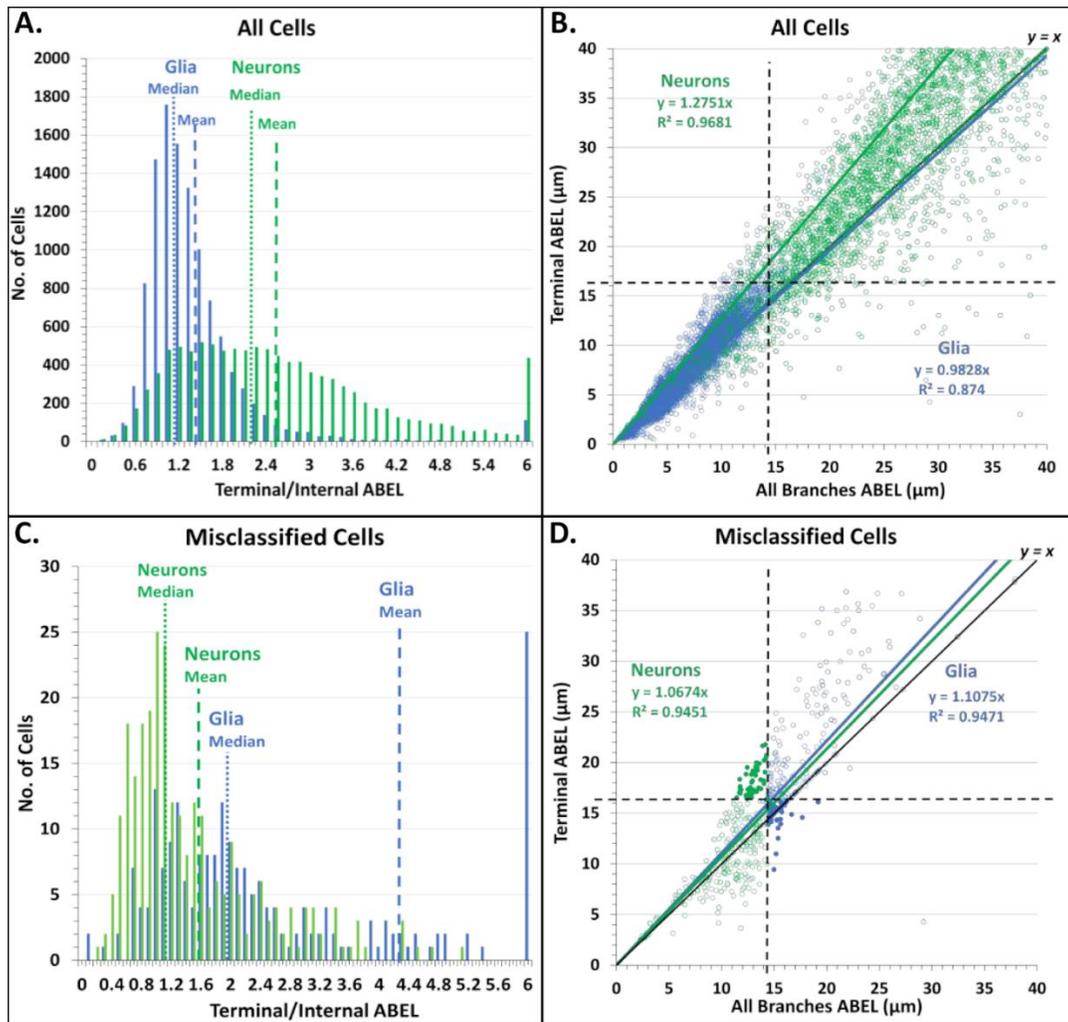


Figure 4. 9. Relationship between the average branch Euclidean length (ABEL) of terminal branches and internal (bifurcating) branches for glia (blue) and neurons (green). (A) Distribution of the ratio between terminal and internal ABEL, with medians (vertical dotted lines) and means (vertical dashed lines) indicated. **(B)** 2D scatter and linear regression between terminal ABEL and all-branch ABEL, with respective classification thresholds indicated by horizontal and vertical dashed lines. **(C)** Same as A except limited to cells that are misclassified based on all-branch ABEL. **(D)** Same as B except limited to cells that

are misclassified based on all-branch ABEL. The filled circles represent the subset of neurons and glia that are misclassified based on all-branch ABEL but correctly classified based on terminal ABEL. An even larger number of cells (not shown) are correctly classified based on all-branch ABEL but misclassified based on terminal ABEL.

Next, we tested the robustness of ABEL as a morphological biomarker of neurons and glia and how well the optimized classification thresholds generalize to new cell datasets. To this aim, we extended the analysis to the additional glial cells released at NeuroMorpho.Org since the beginning of this study and through the time of this writing (v.8.1.90; N=4,286), balancing the dataset with an equivalent number of neurons (N=4,292) from similar species, anatomical regions, and other metadata (as detailed in Supplementary Materials). The ABEL classification accuracy for this new dataset was the same at 97.6% (using the unaltered 14.33 μm threshold). We then tried to assess whether the few outliers were due to systematic patterns or random noise. Classification accuracy was largely consistent across almost all of the metadata investigated,

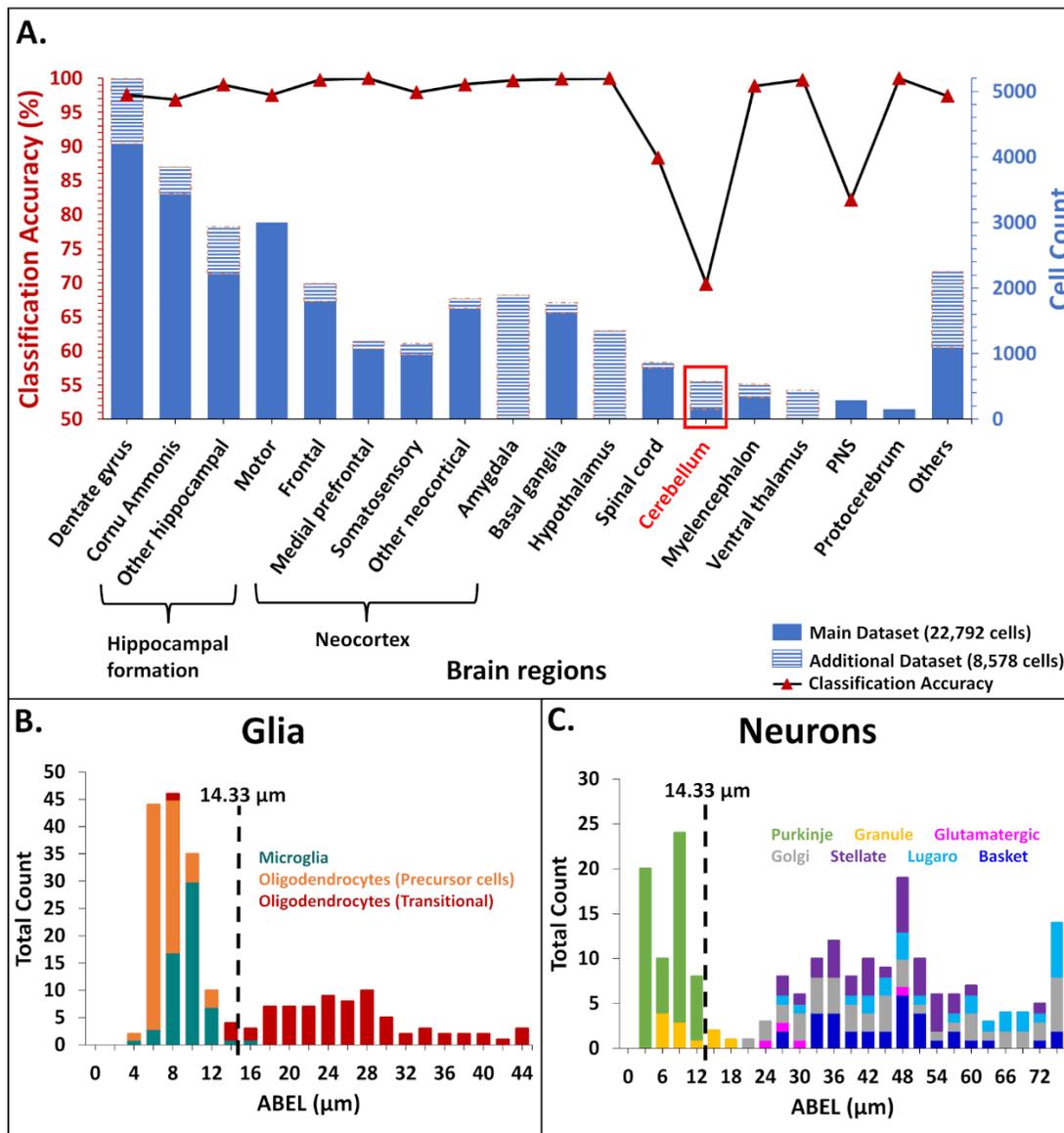


Figure 4.10. Classification of glia and neurons across anatomical regions. (A) Number of cells analyzed (stacked blue bars, right axis: main dataset, solid; and additional dataset, striped) and classification accuracy (black line and red triangle, left axis). (B) ABEL distribution of cerebellar glia. Cells to the right of the threshold (vertical dashed line) are misclassified. (C) ABEL distribution of cerebellar neurons. Cells to the left of the threshold (vertical dashed line) are misclassified.

with only notable exceptions when portioned by brain region (Fig. 10A). Specifically, the high misclassification rate in the cerebellum prompted a deeper evaluation of cells from that region. The misclassified glia consisted of 70 transitional oligodendrocytes and only 1 Iba1-positive microglia, whereas all 78 oligodendrocyte precursor cells, and the rest of cerebellar microglia were correctly classified (Fig. 10B). The 62 misclassified neurons include 8 out of 11 granule cells, and all 54 Purkinje cells, whereas cerebellar basket, stellate, Golgi, Lugaro, and glutamatergic cells were all correctly classified (Fig. 10C). The results indicate that certain cerebellar neurons, specifically Purkinje and granule cells, share similar ABEL with glia. The second, less extreme, exception consisted of the peripheral nervous system (PNS). Here we found that the single culprit was a specific subtype of invertebrate sensory neuron: dendritic arborization (da) Class III cells from the fly larva (46 out of 47 misclassified). In contrast, 104 out of 108 Class I and Class IV sensory neurons, and the quasi totality (98.5%) of PNS glial cells, were correctly classified.

Discussion

Open sharing of digitally reconstructed neuronal morphology from labs across the world has made it possible for researchers to carry out statistical analysis, classification, and computational modeling of their interest (Bota & Swanson, 2007b; Halavi et al., 2012c; Parekh & Ascoli, 2015c). Far fewer morphological classification studies have also included glia, and they typically did not focus on directly comparing neurons to glia. For example, Leyh et al. (2021)

classified different types of microglia in healthy and diseased mouse model, while Zhang et al. (2021) added glia as a separate phenotype in a multiclass neuron type categorization task using convolutional neural networks. Recognizing the morphological signatures that distinguish glia from neurons is an important yet unfulfilled step.

This study sought to determine whether neuronal dendrites and glial processes can be reliably separated solely based on their arbor geometries and independent of animal species, anatomical region, developmental stage, and experimental condition. To this aim, we harnessed all publicly available reconstructions of glia and balanced them with an equivalent number of neurons with as closely matching metadata as possible. The resulting dataset of over 30,000 cells spanned the very broad methodological diversity in the field. We then produced a compact, orthogonalized quantification of those morphologies by applying principal component analysis to an extensive battery of extracted morphometrics. Deployment of three traditional supervised learning algorithms yielded exceptionally high (>99%) classification accuracy. We thus set out to determine which specific differences could explain such striking separation. While neurons were confirmed to have larger arbors than glia, we also discovered that glial trees tend to bifurcate more than neurons, and that glial branches are slightly more tortuous than their neuronal counterparts. These features have already proved useful in the separate investigation of neurons (Kawaguchi et al., 2006; Polavaram et al., 2014), and glia (Khakh & Deneen, 2019; Verkhratsky et al., 2019), but to our best knowledge never in their comparison. Combining these measurements, we defined a novel morphometric parameter, the average branch Euclidean length or ABEL, and demonstrated that it constitutes a powerful and robust morphological biomarker of cell type. Throughout the whole dataset, glia had smaller ABEL values than neurons, and fewer than 2.5% of cells were misclassified based on a simple ABEL threshold of $\sim 14 \mu\text{m}$. Standard

measures of arbor size, such as height, yielded a more than double misclassification rate relative to ABEL.

Molecular expression remains a prominent approach for the consistent identification of cell types in the nervous system. For example, glial fibrillar acidic protein (GFAP), nerve/glia antigen 2 (NG2), and ionized calcium binding adapter molecule 1 (Iba1) are commonly utilized to identify distinct classes of glia. Similarly, neurons are often distinguished by their main neurotransmitter based on presence of vesicular glutamate transporters, glutamic acid decarboxylase, choline acetyl transferase or tyrosine hydrolase. In situ hybridization of the corresponding genes is useful to study the somatic distribution of these neurons and glia but does not reveal their dendrites and processes. Immunolabeling can in some cases visualize cell type-specific neural arbors, and multi-color combinations of antibodies may allow co-labeling of distinct cell types in the same preparation. In contrast, relatively simpler but non-selective staining such as Golgi (Ghosh, 2020) impregnates a broad spectrum of neurons and glia. In these cases, ABEL can provide a practical way to quickly recognize neurons from glia. It is important to note in this regard that measuring ABEL does not necessarily require the detailed tracing of the full arbors. Euclidean length is simply defined as the straight-line distance between the start and end points of a branch, which can be computed directly from the microscopic image in any common software. Moreover, we showed that as few as five branches are sufficient to provide an ABEL approximation that distinguishes glia from

neurons with >95% accuracy. Even for complex arbors with hundreds of bifurcations and terminations, it is thus possible to estimate ABEL with minimum effort.

Besides the practical utility, it is tempting to speculate about the possible scientific interpretation of our main finding. The systematically small ABEL values of glia suggest a tendency to optimize spatial occupancy, consistent with extensively reported tiling properties for these cells (Barber et al., 2021; Pogodalla et al., 2022). In contrast, the larger ABEL values of neurons are indicative of pressure to maximize spatial exploration, in line with the role of dendrites to integrate converging synaptic signals from multiple neural pathways (Anton-Sanchez et al., 2018; Stepanyants & Chklovskii, 2005). It is especially intriguing to consider the rare exceptions that emerged from our analysis. Since the only glial outliers in terms of ABEL were transitional oligodendrocytes, it is possible that the compact arbor is an acquired property of mature glia rather than an innate feature, and that seeking myelination targets requires a degree of spatial exploration. The main neuronal exceptions were cerebellar granule and Purkinje cells. It may not be a coincidence that these two neuron types together form one of the most peculiar circuits in any neural system: the parallel fibers of the cerebellum, which ascend from granule cell axons and contact the Purkinje dendrites on up to 100,000 spines. Purkinje cells are the output cells of the cerebellar cortex, and their dense, planar dendrites are fan-shaped and branch extensively to cover the field of their respective territories without overlapping (Fujishima et al., 2018). These features, which push Purkinje dendrites towards the compact spatial occupancy typical of glia, are

dictated by the need to sample the exceptionally large number of synaptic signals from the parallel fibers (Hirano, 2018). Cerebellar granule cells are the single most abundant neuron type in the mammalian brain (Herculano-Houzel, 2010) as well as the most densely packed (Badura & De Zeeuw, 2017), leading to considerably small dendritic fields (Houston et al., 2017). These characteristics, again determined by the unique connectivity profile of the cerebellar parallel fibers, are more akin to those of glial processes than of typical neuronal dendrites. Of note, the other cerebellar neurons (basket, Lugaro, Golgi, and stellate cells) are all correctly classified by ABEL. These exceptions point to a clearly different cell organization in the cerebellum compared to other brain regions.

Aside from the sparse exceptions, the robustness of the results reported in this study is underscored by the very large dataset, distributed provenance of the reconstructions, and broad diversity of metadata. At the same time, it is also essential to recognize that this study is intrinsically limited by the data availability. For example, although the included species span primates, rodents, fish, and invertebrates, the majority of reconstructions for both neurons and glia come from rats and mice. Furthermore, while many anatomical regions are represented in the study, the list is far from complete. And albeit several classes of glia and of neurons were analyzed, their distribution was far from uniform. These factors reflect the state of the research in this field rather than a flawed analysis design. Nevertheless, the conclusions must be considered tentative until further validated as more data continue to accumulate.

This work also illustrates the usefulness of subjecting very large datasets to exploratory analysis via machine learning, followed by a targeted investigation of the most promising phenomena or patterns revealed. This “breadth-then-depth” approach may help shed light on otherwise elusive mechanisms. In particular, tracing glial morphology has become progressively more common and, thanks to increased awareness of data sharing, ever larger amounts of glial reconstructions are being deposited to NeuroMorpho.Org. This increment in data availability in a public repository opens new doors for scientific discovery, especially when applying different analysis and modeling techniques for glia that have been productively applied to neurons since the early days of computational neuroscience.

Supplementary material

The following files are available at https://github.com/Masood-Akram/Classification_Neurons-Glia/tree/main/Supplementary_Material

Scale Correction Main Dataset: calculations of the correction factors for the archives of the main dataset reporting reconstruction coordinates in pixels rather than microns.

Scale Correction Additional Dataset: calculations of the correction factors for the archives of the additional dataset reporting reconstruction coordinates in pixels rather than microns.

Metadata Dimensions Main Analysis: detailed breakdown of the metadata for all archives of the main dataset.

Metadata Dimensions Additional Analysis: detailed breakdown of the metadata for all archives of the additional dataset.

APPENDIX 1: GENETIC SINGLE NEURON ANATOMY REVEALS FINE GRANULARITY OF CORTICAL AXO-AXONIC CELLS

Xiaojun Wang^{1,2}, Jason Tucciarone^{3,4,7}, Siqu Jiang^{1,2,7}, Fangfang Yin^{1,2}, Bor-Shuen Wang³, Dingkan Wang⁵, Yao Jia^{1,2}, Xueyan Jia^{1,2}, Yuxin Li^{1,2}, Tao Yang^{1,2}, Zhengchao Xu^{1,2}, Masood A. Akram⁶, Yusu Wang⁵, Shaoqun Zeng^{1,2}, Giorgio A. Ascoli⁶, Partha Mitra³, Hui Gong^{1,2}, Qingming Luo^{1,2*}, and Z. Josh Huang^{3,8*}

¹Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics-Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

²MoE Key Laboratory for Biomedical Photonics, Collaborative Innovation Center for Biomedical Engineering, School of Engineering Sciences, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

³Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

⁴Program in Neuroscience and Medical Scientist Training Program, Stony Brook University, Stony Brook, NY 11790, USA

⁵Computer Science and Engineering Department, The Ohio State University, Columbus, OH 43221, USA

⁶Center for Neural Informatics, Structures, and Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA

⁷These authors contributed equally

*This appendix was published in Cell Reports, 26(11), 3145-3159.e5 (2019).

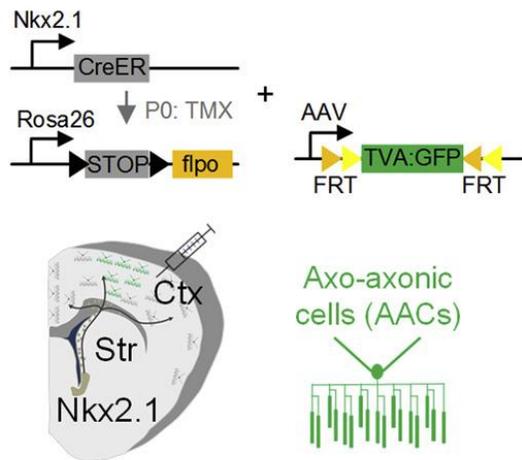
Summary

Parsing diverse nerve cells into biological types is necessary for understanding neural circuit organization. Morphology is an intuitive criterion for neuronal classification and a proxy of connectivity, but morphological diversity and variability often preclude resolving the granularity of neuron types. Combining genetic labeling with high-resolution, large-volume light microscopy, we established a single neuron anatomy platform that resolves, registers, and quantifies complete neuron morphologies in the

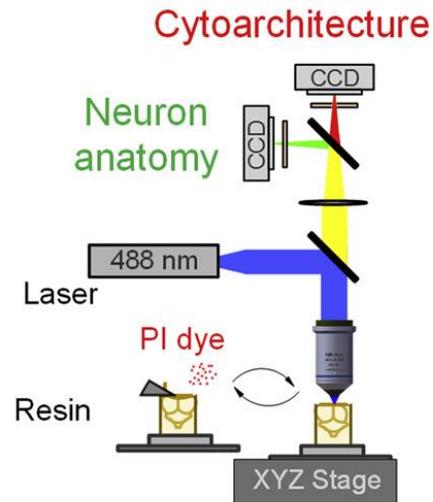
mouse brain. We discovered that cortical axo-axonic cells (AACs), a cardinal GABAergic interneuron type that controls pyramidal neuron (PyN) spiking at axon initial segments, consist of multiple subtypes distinguished by highly laminar-specific soma position and dendritic and axonal arborization patterns. Whereas the laminar arrangements of AAC dendrites reflect differential recruitment by input streams, the laminar distribution and local geometry of AAC axons enable differential innervation of PyN ensembles. This platform will facilitate genetically targeted, high-resolution, and scalable single neuron anatomy in the mouse brain.

Graphical Abstract

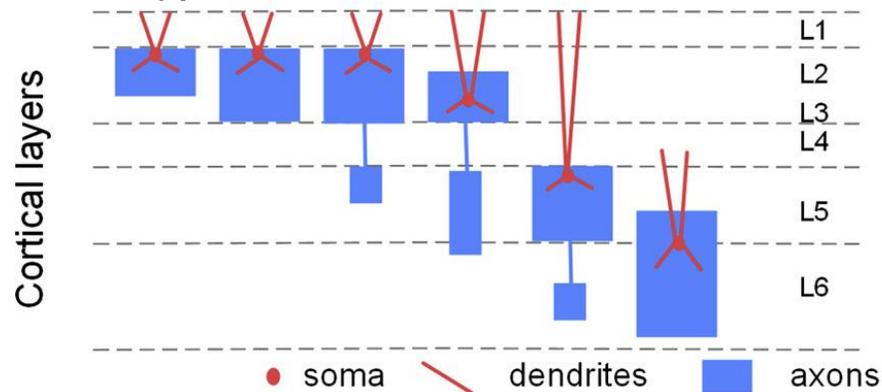
1. Genetic viral labeling



2. dfMOST imaging



3. AAC subtypes



Introduction

Defining and cataloging neuronal cell types, groups of neurons that share anatomical, physiological, and molecular properties are necessary for understanding the organizational logic of neural circuits (Huang & Zeng, 2013). As phenotypic variations of neurons often span substantial parameter space, it is necessary to carry out

comprehensive, quantitative, and scalable single-cell analysis to resolve the appropriate granularity of cell type definition (Zeng & Sanes, 2017). Recent advances in single-cell RNA sequencing (scRNA-seq) enable quantitative measurements of cellular transcriptome profiles at massive scale, and computational analyses reveal increasing number of “transcriptional types” and discrete as well as continuous variations (Macosko et al., 2015; Tasic et al., 2018; Zeisel et al., 2015). As neuronal phenotypes are inherently multi-modal, it is necessary to achieve single-cell analyses of orthogonal cell features toward an integrated definition of neuron types that encapsulate the issue of granularity.

Neuronal morphology has been an intuitive first-level description of cell types. In several invertebrate systems (Aso et al., 2014; Chiang et al., 2011) and the vertebrate retina (Sanes & Masland, 2015) in which neurons are relatively small and stereotyped, comprehensive and quantitative single-neuron morphometry has allowed operational and consensual definition of neuron types. Morphology-based cell catalogs in these systems have been achieved or within reach (Aso et al., 2014; Hobert et al., 2016; Seung & Sümbül, 2014), which provide a foundation for multi-modal analysis and for exploring neural circuit organization. In the mammalian brain, however, the vast diversity, large spatial span, and seemingly endless variations of neuronal shapes present unique challenges in morphological tracing and analysis (Huang & Zeng, 2013; Lichtman & Denk, 2011). Single-neuron anatomy in the mammalian brain requires overcoming several technical hurdles. The first is labeling: to systematically, reliably, sparsely, and completely label specific sets of individual neurons. The second is imaging: to achieve

axon resolution imaging in brain-wide volume (Economo et al., 2016; Gong et al., 2013; A. Li et al., 2010). The third is cell reconstruction: to convert large image stacks into digital datasets of single-neuron morphology. The fourth is analysis: to register neuronal morphology within appropriate spatial coordinate framework, and to extract, quantify, and classify biologically relevant attributes (e.g., those relate to neural connectivity).

Here, we present a robust genetic single neuron anatomy (gSNA) platform in the mouse that overcomes some of these challenges. We combined genetic cell labeling with dual-color fluorescence micro-optical sectioning tomography (dfMOST) (Gong et al., 2016) to achieve axon resolution and brain-wide imaging and spatial registration of genetically targeted single neurons. We focused our analysis on one of the most distinctive cortical GABAergic interneurons: the axo-axonic cells (AACs) that specifically innervate the axon initial segment (AIS) of glutamatergic pyramidal neurons (PyNs) (Somogyi et al., 1982; Taniguchi et al., 2013; Woodruff et al., 2010) and likely control spike initiation. Complete reconstruction of single AACs and their precise registration along cortical laminar coordinate allowed quantitative morphological analysis in the context of input-output connectivity. We discovered that cardinal AACs consist of multiple discrete subtypes that are distinguished by highly laminar-specific soma position and dendritic and axonal arborization patterns. The laminar arrangements of AAC dendrites may allow differential recruitment by presynaptic input streams. Furthermore, the laminar stratification of AAC axon arbors correlates with the distribution of PyN subsets and the local geometry of AAC axon terminals differentially conform to the laminar features of PyN AIS, suggesting differential innervation of PyN ensembles. Our

results support a hierarchical scheme of neuronal classification (Zeng & Sanes, 2017) and suggest that cardinal neuron types consist of fine-grained subtypes, which can be deduced from light microscopy and mesoscale analyses that inform input-output connectivity patterns. The gSNA platform enables scalable and comprehensive single-neuron anatomical analysis, which will provide foundational datasets for neuron type discovery and classification in the mammalian brain.

Results

Establishing a gSNA Platform

Our gSNA platform consists of four components (Figure A1.1). The first is a method to systematically label different sets of genetically targeted individual neurons to their entirety; the second is a technology for simultaneous imaging of labeled neurons at axon resolution and all other cell body positions throughout the entire mouse brain (dfMOST) (Gong et al., 2016); the third is a procedure to completely reconstruct single neurons from brain volume image stacks; the fourth is an analysis pipeline that registers and quantifies neuronal morphology within an appropriate spatial coordinate system that reflect network connectivity. Here, we integrate genetic labeling with fMOST in the gSNA platform to analyze the morphological diversity of a well-recognized interneuron type in the cerebral cortex.

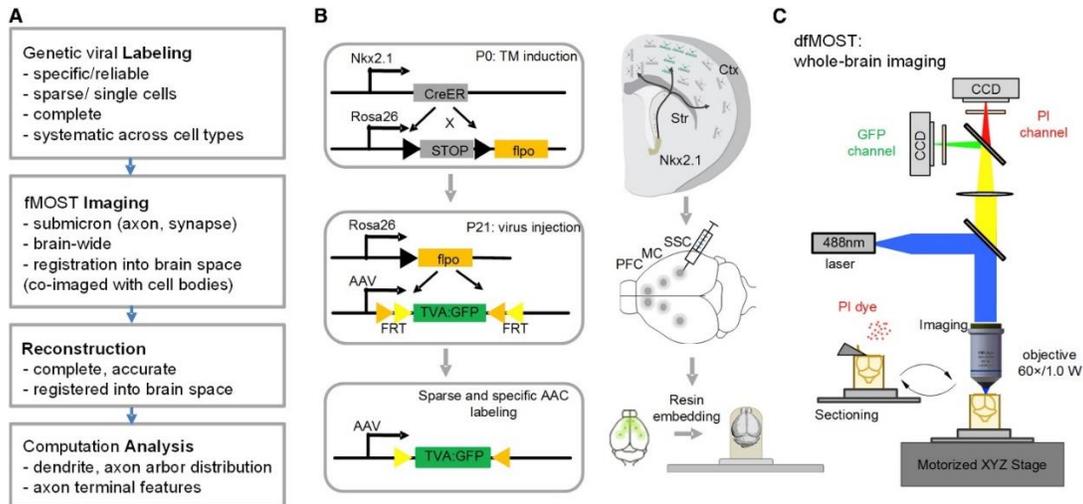


Figure A1.1. Schematic of the gSNA Platform Applied to the Mouse Brain. (A) Pipeline and components of genetic single neuron anatomy (gSNA). (B) Left: scheme of genetic and viral strategy for the labeling of axo-axonic cells (AACs). A transient CreER activity in MGE progenitors is converted to a constitutive Flp activity in mature AACs. Flp-dependent AAVs injected in specific cortical areas enables sparse and robust AAC labeling. (C) fMOST high-resolution whole-brain imaging. Two-color imaging for the acquisition of GFP (green) channel and PI (propidium iodide, red) channel signals. PI stains brain cytoarchitecture in real time, and therefore provides each dataset with a self-registered atlas. A 488-nm wavelength laser was used for the excitation of both GFP and PI signals. Whole-brain coronal image stacks were obtained by sectioning (with a diamond knife) and imaging cycles at 1- μ m z steps, guided by a motorized precision XYZ stage.

AACs were recognized as a bona fide type largely based on their unique morphology and specific innervation of PyNs at AIS (Somogyi et al., 1982). Although the precise physiological action of AACs remains to be elucidated (J. Lu et al., 2017; Szabadics et al., 2006; Woodruff et al., 2010), the defining feature is their specialization in regulating the spike initiation of PyNs. As multiple morphological variants of AACs have been found in several cortical structures (e.g., the hippocampus [(Szabo et al., 2017;

Varga et al., 2014)], piriform cortex, and neocortex) and in different cortical layers (Defelipe et al., 1985; Lewis & Lund, 1990; Somogyi et al., 1982; Taniguchi et al., 2013), this raises the questions of whether the cardinal AAC type consist of multiple “subtypes,” and how AAC subgroups should be defined. We have previously captured cortical AACs through genetic fate mapping of neural progenitors of the embryonic medial ganglionic eminence (MGE) using the *Nkx2.1-CreER* driver line (Taniguchi et al., 2013). Conversion of transient Nkx2.1-CreER expression in MGE progenitors to a constitutive Flpase activity in AACs enabled postnatal viral targeting (He et al., 2016). By controlling CreER efficiency (i.e., tamoxifen dose) and AAV injection volume and location, we were able to achieve specific, sparse, and complete labeling of AACs in defined cortical areas (Figures A1.2 and A1S1). Here, we analyzed AACs in the medial prefrontal (mPFC), motor (MC), and somatosensory (SSC) cortex. We use the original nomenclature *axo-axonic cells* (AACs) (Somogyi et al., 1982) to refer to all GABAergic interneurons that innervate PyNs at AIS. Under this category, we use the term *chandelier cells* (ChCs) to refer to the subsets of AACs in the cerebral cortex (especially those in supragranular layers), whose axon arbors resemble the candlesticks of a chandelier light.

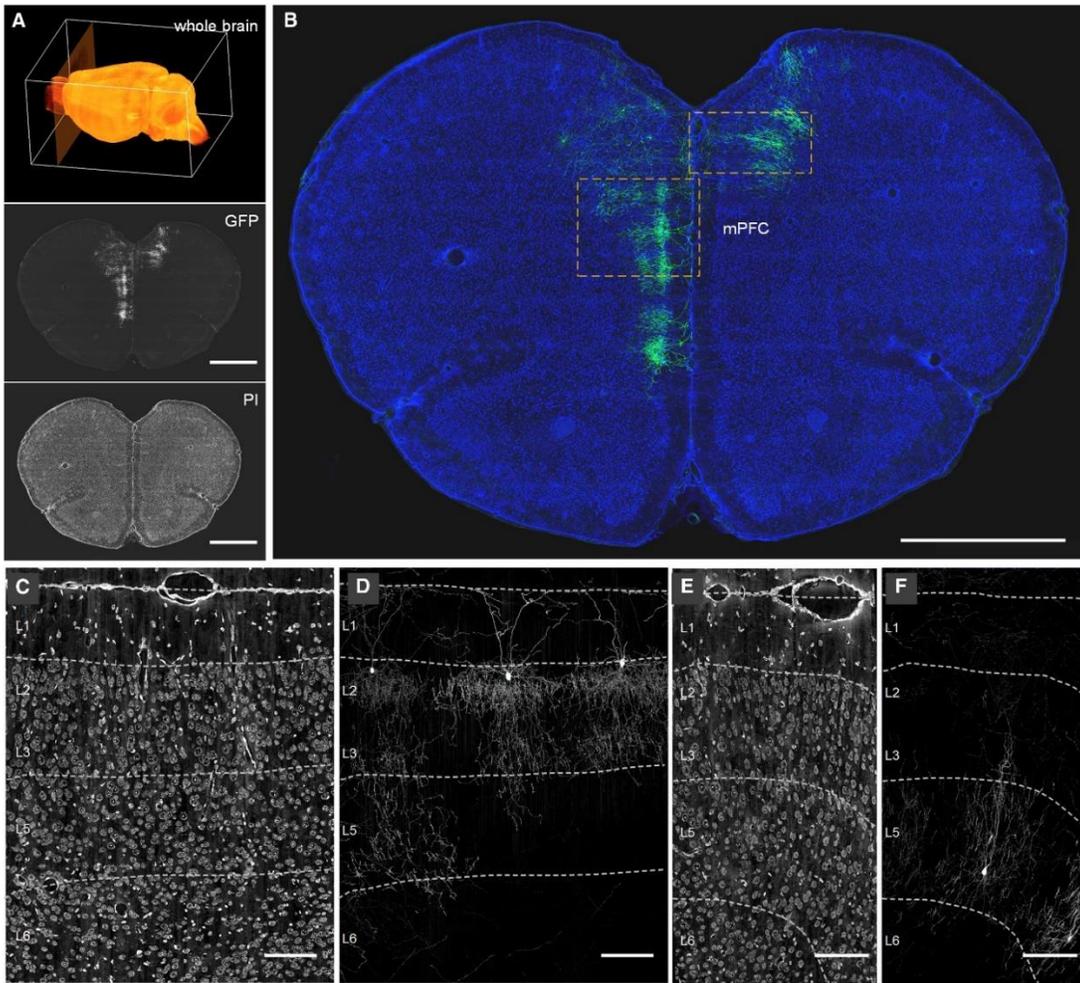


Figure A1. 2. Areal and Laminar Distribution of AACs Revealed from Whole-Brain fMOST Dataset.

(A) A schematic of whole-brain coronal dataset collection (top) with an example of GFP channel (center) and PI channel (bottom) images. Scale bars: 1,000 μm . (B) An example of the distribution of sparsely labeled AACs in mPFC. Green: AAC morphology, 100- μm max-intensity projection. Blue: cytoarchitecture revealed by PI, 5- μm max-intensity projection. Scale bar: 1,000 μm . (C and D) Laminar distribution of L2 AACs in mPFC. Enlargement of PI channel (C) and GFP channel (D) images from the left box in (B). (E and F) Laminar distribution of L5 AACs in mPFC. Enlargement of PI channel (E) and GFP channel (F) images from the right box in (B). Dashed lines in (C)–(F) indicate the layer boundaries. Cortical layers were discriminated based on cell body distributions in PI channel according to the Allen Mouse Brain Reference Atlas (<http://portal.brain-map.org/>). Scale bars in (E) and (F): 100 μm .

Following tissue resin embedding and processing (Gong et al., 2016; Xiong et al., 2014), we used a dfMOST system to image the whole-brain samples at submicron resolution (Figure A1.1C; Video A1S1). The dual-channel capturing of neural morphology labeled by GFP and brain cytoarchitecture stained by propidium iodide (PI) (red) were achieved by using a wide-field upright epi-fluorescence microscopy with a blue laser (488 nm) for fluorescence excitation and two separate charge-coupled device (CCD) cameras for signal detection (details in STAR Methods). Importantly, the PI channel provided each brain dataset with a self-registered Nissl-like reference atlas of cell body distribution information, which allowed reliable delineation of cortical areas and layer boundaries (Figures A1.2, A1.3A, and A1S1). Furthermore, the image contrast in PI channel was sufficient for the reconstruction of PyN main dendrites, which were used for identifying local laminar and vertical coordinates, readjusting cell orientation, and establishing a standardized platform for comparative analysis between cells in different cortical areas (Figures A1.3A and A1.3B).

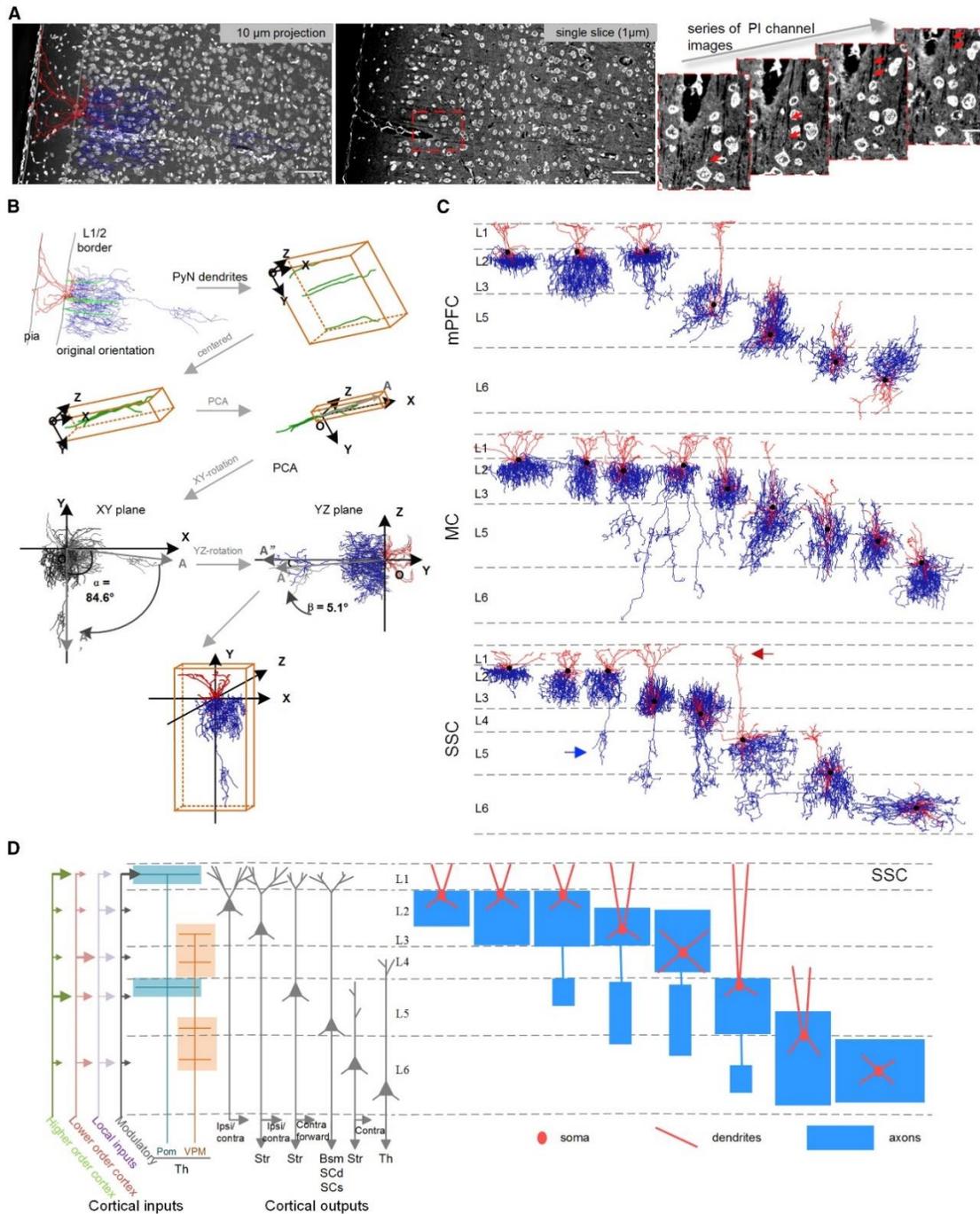
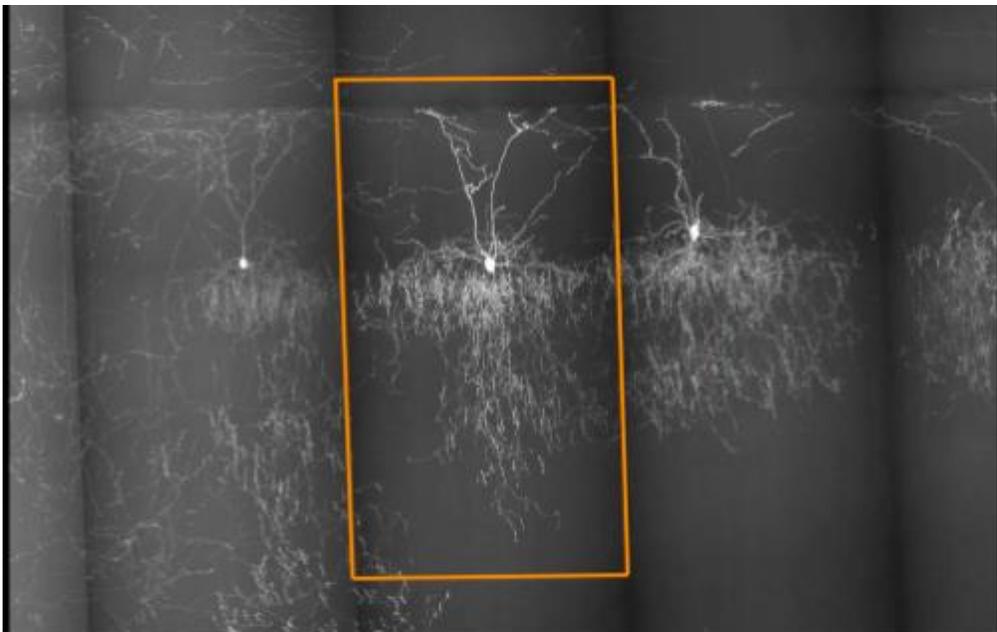


Figure A1. 3. Single AAC Reconstructions in Cerebral Cortex. (A) A representative AAC reconstruction and its co-registered PI channel images. Left: overlay of the reconstructed cell with its PI channel image (10- μ m max intensity projection) with the original orientation. Cytoarchitecture information shows the cortical laminar organization (more details in [STAR Methods](#)). Scale bar: 50 μ m. Middle: single slice of PI channel. Right: enlarged image series from the boxed area in middle panel. Arrows indicate a pyramidal neuron main dendrite extending from cell body. Scale bar: 15 μ m. (B) Rotation and alignment

procedures based on reconstructed pyramidal dendrites (green). Randomly selected pyramidal dendrites near the AAC cell body were reconstructed in NeuroLucida360. The vertical axis of the local cortical column was calculated by performing principal-component analysis (PCA) on the centered dendritic reconstructions. AAC reconstruction was then re-aligned in the coronal plane (XY plane) and sagittal plane (YZ plane) around the cell body based on the identified cortical column orientation. (C) Representative AAC single-cell reconstructions in mPFC, MC, and SSC. Cortical layers in each area were indicated by dashed lines. Black: soma body. Red: dendrites. Blue: axons. The orientation of each reconstruction was re-adjusted according to the local cortical vertical axis (see more details in [STAR Methods](#)). Representative translaminal axonal and dendritic arbors in the SSC are indicated by blue and red arrows, respectively. (D) Left: scheme of the laminar arrangement of the input and output streams of SSC, in part rooted in the laminar organization of pyramidal neuron types with distinct projection targets. Right: a schematic of representative AACs in the SSC with characteristic laminar dendritic and axonal distribution patterns. Str, striatum; Bsm, brainstem; Scd, spinal cord; SCs, superior colliculus; Pom, posterior complex of thalamus; VPM, ventral posteromedial nucleus of the thalamus; Th, thalamus; ipsi, ipsilateral; contra, contralateral.



[Download video](#)

Supplementary Video A1.1. dfMOST Imaging of Viral-Labeled AACs at Single-Axon Resolution, Related to Figure A1.1. 100 μm max intensity projection of original GFP channel images without any imaging processing.

From 11 whole-brain dfMOST datasets, we completely reconstructed 62 AACs from mPFC, MC, and SSC (Figures A1.3C and A1S2A; Table A1S1). As axon arbors of AACs were extremely dense and complex, all AACs were manually reconstructed. With rare exceptions (Bienvenu et al., 2012; Viney et al., 2013), previous labeling of AACs were largely carried out in brain slice preparations where axons and dendrites were severed, and thus reconstructions were mostly incomplete (Blazquez-Llorca et al., 2015; Kawaguchi & Shindou, 1998; Somogyi et al., 1982; Woodruff et al., 2010). Our dataset represents a set of complete and comprehensive AAC reconstructions in the cortex since their discovery four decades ago (Szentágothai & Arbib, 1974). The average length of AAC axons was 2.14 ± 0.79 cm ($n = 62$; mean \pm SD), average number of axon branches was $1,369 \pm 499$ ($n = 62$; mean \pm SD), and average axon branch order was 32 ± 8 ($n = 62$; mean \pm SD). A major goal of our analysis is to define and discover AAC subtypes based on morphological features that inform connectivity, taking full advantage of the obligatory synaptic relationship between AAC axon terminals and PyN AIS. Our strategy was to examine the location and distribution of AAC cell bodies, their dendrite and axon arbor distribution, and their axon arbor geometry in the well-established coordinates of cortical laminar organization based on AAC postsynaptic targets—the PyNs (Figure A1.3D).

AACs Tend to Localize at the Borders between Cortical Layers

Previous studies in several species found that AACs are distributed across most if not all cortical areas and in multiple cortical layers (Lewis & Lund, 1990; Somogyi et al., 1982; Taniguchi et al., 2013), but more precise description of AAC distribution and

positioning has not been reported. The cellular resolution spatial coordinate information in the dfMOST datasets allowed unambiguous and quantitative localization of AACs. Within all three areas, the largest proportion of our reconstructed AACs was located within the supragranular layers, with a major fraction at the layer 2 to layer 1 (L2/1) border (55%) and a much smaller set in L3 (5%) ([Figures A1.2](#), [A1.3C](#), and [A1S1](#)). A significant portion of AACs were found in infragranular layers, both in L5 (L5 22%) and L6 (16%). We found one AAC in L4 of SSC in our dataset. Interestingly, in most cases, AAC somata tended to localize at the border between cortical layers, with prominent apical dendrites and basal axons ([Figures A1.2](#), [A1.3C](#), and [A1.3D](#)).

AACs Elaborate Laminar-Restricted Apical and Basal Dendrites That Protrude Dendritic Spines

Almost all of the reconstructed AACs elaborated prominent apical dendrites ([Figure A1.4](#)). The average span of apical dendrites of L2 AACs was 85.0 μm (90% of dendrite arbors horizontally cover $85.0 \pm 23.0\text{-}\mu\text{m}$ radial distance; mean \pm SD; $n = 61$) from soma ([Figures A1.4G](#) and [A1S3](#)). In most cases (51/62; 82% of all reconstructed AACs), the apical dendrite extended within the one layer above the soma location (e.g., L1 for L2 AACs and L5 for L6 AACs). In several cases, L3 (3 cells) and L5 (4 cells) AACs extended apical dendrites all the way to the pia ([Figure A1.3C](#); [Videos A1S4](#) and [A1S5](#)). In particular, all L2 (34 AACs) and some L3 (3 cells), L5 (2 cells) AAC dendrites appeared to tightly attach to the pia with thickened apical tufts; this is in contrast to many PyN apical dendrites in L1 that do not reach near or adhere to pia

surface (Figures A1.4A, A1.4B, A1.4F, and A1S3B). Interestingly, the apical but not basal dendrites of L2 ChCs sprouted filopodia-like slender dendritic spines, which were enriched in the upper half (68% in upper L1, the rest near L1/2 border) of L1 (Figures A1.4C–A1.4K).

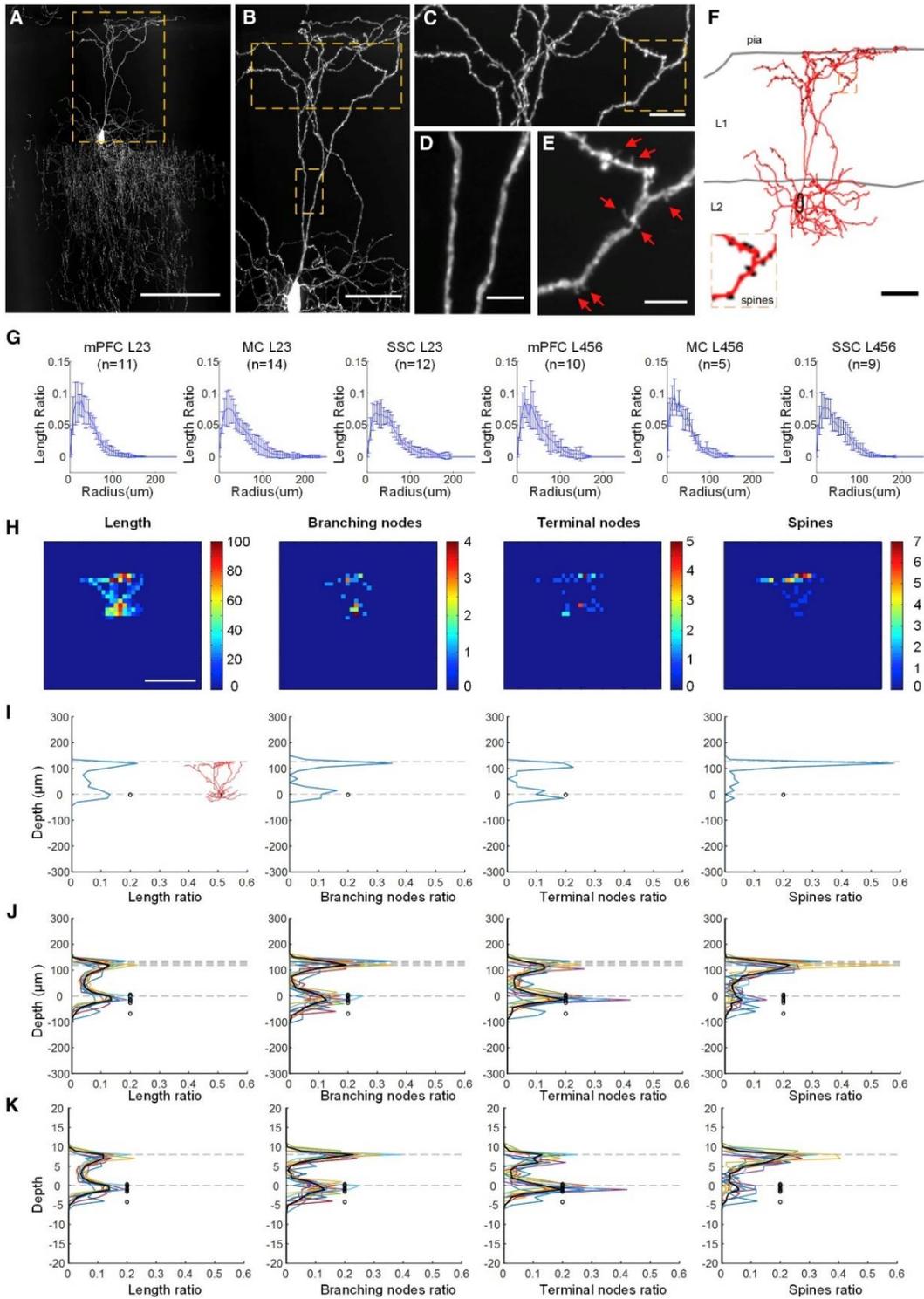
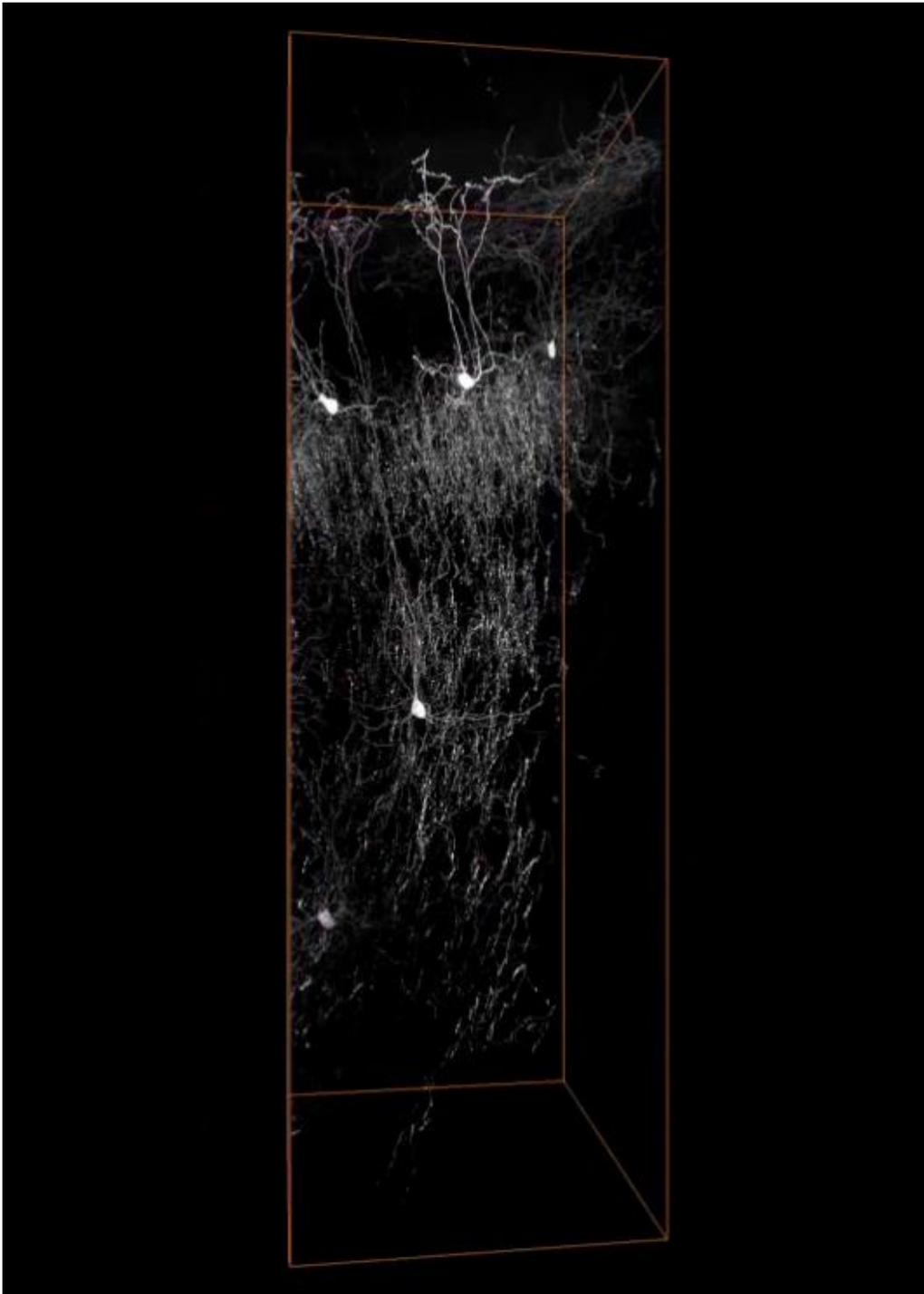


Figure A1.4. Characteristics of L2 AAC Dendrites. (A) A representative L2 AAC in mPFC. 100- μ m max intensity projection. Scale bar: 100 μ m. (B) Dendrites of L2 AAC. Image was enlarged from the box

in (A). Scale bar: 50 μm . (C and D) Apical (C) and main dendrites (D) were enlarged from boxes in (B). Scale bars: 30 μm (C) and 5 μm (D). (E) Spines (arrows) on the apical dendrites were enlarged from the box in (C). Scale bar: 5 μm . (F) Complete reconstruction of dendrites (red) and spines (black). The same cell shown in (A) and (B). Inset: enlarged from the box. Black circle: cell body. Gray lines: pia and L1/2 border. Scale bar: 50 μm . (G) Horizontal dendritic arbor distributions of up-layer (L2 and L3) and deep-layer (L4, L5 and L6) AACs in mPFC, MC, and SSC. Data are mean \pm SD. (H) An example of heatmaps showing the density distribution patterns of a L2 AAC dendritic arbor length (left), branching nodes (middle left), terminal nodes (middle right), and spines (right). Scale bar: 200 μm . (I) Single-cell density plots of L2 AAC dendrites (same as in H) along the cortical depth. (J) Density plots of dendrites from 11 L2 AACs in mPFC. Different colors indicate different cells.

(K) Normalized density plots of (J) based on pia and L1/2 border positions (see more details in [STAR Methods](#)). Black circles in (I)–(K) indicate AAC soma positions in the coordinate. Dashed lines correspond to the place of pia (top) and L1/2 border (bottom). Dark black curves in (J) and (K) are averages of all the cells. Density value was presented by ratio.



[Download video](#)

Supplementary Video A1.5. A L5a (L5-cross) AAC in mPFC, Related to Figure A1.3.

Although overall more sparse than apical dendrites, the basal dendrites of AACs show striking laminar restriction to the same layer of the cell soma (Figures A1.4H–A1.4K, A1S3C, and A1S3D). For example, all L2 AACs (n = 34) restrict their basal dendrites strictly to L2 without extending to L3, while L3 AACs (n = 3) restrict their basal dendrites strictly to L3 without extending to L2 and L4 (Figures A1S2A and A1S5D). Similarly, L5 and L6 AAC basal dendrites manifest the same intralaminar restriction. Together, these results suggest that AACs elaborate their dendritic arbors in a laminar-specific and unitary pattern instead of a continuous and diffuse pattern. The polarized dendritic arborization suggests that AACs receive most of their inputs from above their cell bodies; in particular, pia-attached AAC dendrites may recruit the most superficial L1 inputs and select or modify these inputs through dendritic spines. On the other hand, the basal dendrites receive inputs strictly targeting the same layer where the cell soma is located.

AACs Elaborate Laminar-Stratified Axon Arbors, Some with Translaminar Arbors

Although the characteristic shape and exquisite specificity of AAC axons have been recognized decades ago, few if any have been reconstructed to their entirety. We found that AACs axons arborized very extensively near the cell soma (below the soma for L2 AACs and both above and below the soma for other cortical AACs; Figures A1.5A–A1.5C, A1S4, and A1S5A). The average span of AAC axon arbors was 129.2 μm

(90% of axon arbors horizontally cover 129.2 ± 27.5 - μm radial distance; mean \pm SD; n = 61). In addition to the highly predominant local arbor (i.e., intralaminar; [Figure A1.5A](#)), a significant fraction of L2 and L3 AAC axons ($\sim 74\%$ of our L2 AAC reconstructions) further extended to the deeper layers (i.e., cross- and trans- laminar; [Figures A1.3C](#), [A1.3D](#), [A1.5A](#), [A1.5B](#), [A1.5D](#), [A1S2](#), [A1S5A](#), [A1S5D](#), and [A1S5F](#); [Videos S2](#) and [S3](#)). In particular, translaminar axons of L2 AACs descended through intervening layers (e.g., L3–L5A in MC or L4 in SSC) before elaborating terminal branches with presynaptic boutons ([Figure A1.5D](#)). This result suggests that, in addition to exerting powerful control over local PyN populations, some L2/3 AACs likely coordinate firing between local PyNs and a distant ensemble in an infragranular layer. Overall, AAC axon arbors appear to strictly conform to laminar borders instead of extending diffusely across layers, suggesting a laminar-specific and unitary pattern of axon stratification. Interestingly, we observed one L6 AAC with an inverted polarity—its dendrite extended below toward the white matter, whereas the axon extended above toward L5 ([Figures A1.3C](#), [A1S5E](#), and [A1S5G](#)).

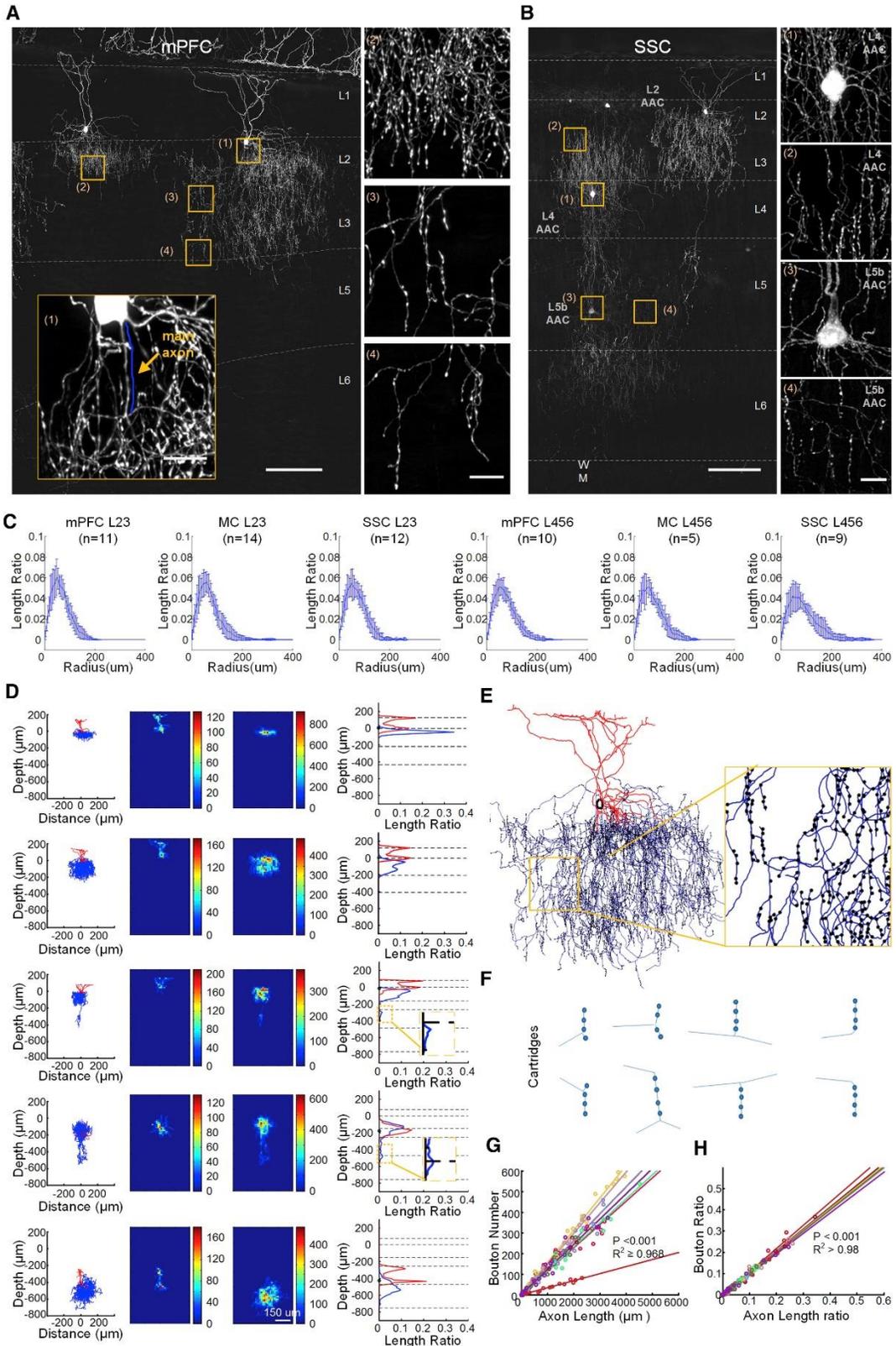
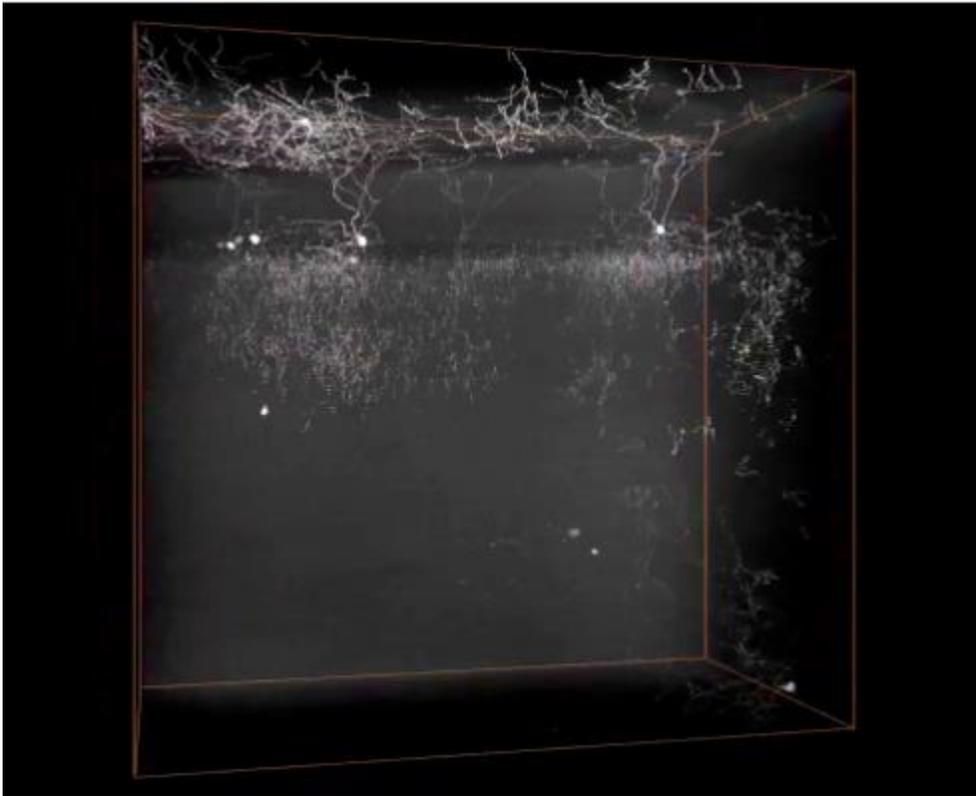
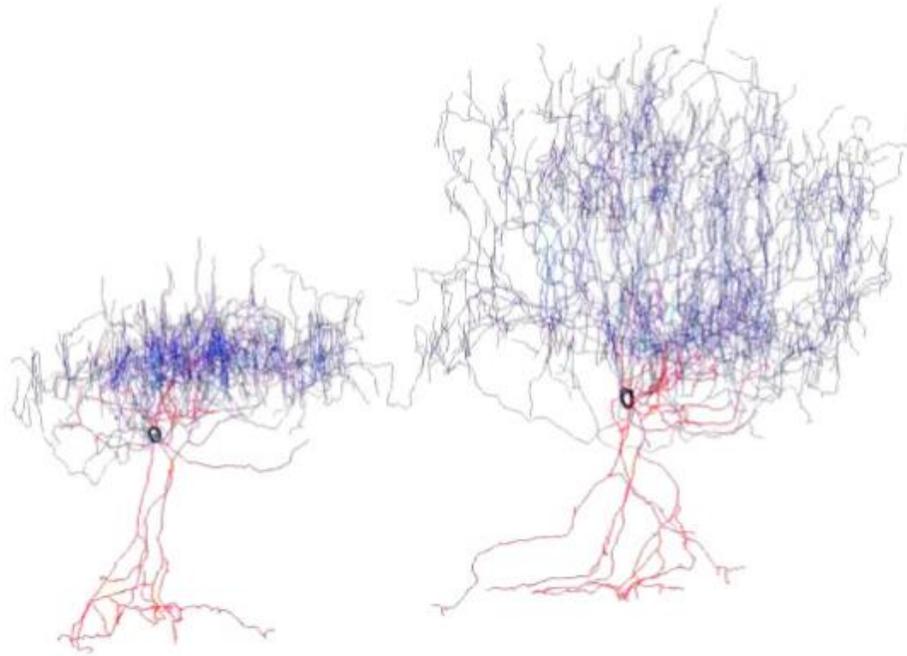


Figure A1. 5. Morphology and Distribution Patterns of AAC Axons in the Neocortex. (A) A representative image of two examples of nearby intra- (left) and cross- (right) L2 AACs in mPFC. Insets are enlarged images from boxed regions showing the main axon extending from the soma (1; arrow), characteristic axon cartridge clusters, and individual boutons from different regions of the axon arbor (2, 3, and 4). Image is a projection of 100- μ m image stack. Scale bars: 100 μ m (low-mag image) and 10 μ m (insets). (B) A representative image containing nearby L2, L4, and L5 AACs in SSC. Enlarged L4 and L5 AACs were from the boxes in the left panel. Scale bars: 100 μ m (left) and 10 μ m (right). Dashed lines in (A) and (B) indicate cortical layers. (C) Horizontal axon arbor distributions of up-layer (L2 and L3) and deep-layer (L4, L5 and L6) AACs in mPFC, MC, and SSC. Data are mean \pm SD. (D) Length density analysis of axons and dendrites from all the AACs shown in (A) and (B). Left: projection of reconstructions (dendrites in red; axons in blue). Middle: heatmap of length density distribution of dendrites (middle left) and axons (middle right). Right: length density plots of AAC dendrites and axons along cortical depth (y axis). Dashed lines indicate layer boundaries. Insets in rows 3 and 4 highlight axon branches in deep layers. (E) An example of axon bouton reconstruction of L2 AAC in mPFC. Inset: magnified view of the boxed region. (F) Axon cartridges that innervate PyN AIS can point upward, downward, or split from the middle. (G and H) The numbers of synaptic boutons correlate with axon length quantified by absolute value (G) or ratio (H).



[Download video](#)

Supplementary Video A1.2. Nearby L2-Intra and L2-Cross AACs in mPFC, Related to Figure A1.3.



[Download video](#)

Supplementary Video A1.3. Reconstructions of Nearby L2-Intra and L2-Cross AACs in mPFC, Related to Figure A1.3.

AACs Consist of Multiple Subtypes Distinguished by Dendrite-Axon Distributions That Reflect Input-Output Connectivity Patterns

The substantial variations in the location and morphology of AACs raise questions of whether they consist of anatomical “subtypes” and how subtypes can be resolved with biologically relevant properties. As morphology is a proxy to and serves the purpose of connectivity, we first adopted a connectivity-guided approach to morphology-based AAC subtyping. Our analysis was based on the premise that, at a mesoscale, establishing a synaptic connection requires the physical overlap between a

presynaptic axon and its postsynaptic element within a specific anatomic location, i.e., an “anatomic parcel,” that represents the input or output component of a neural network (Ascoli & Wheeler, 2016); this tight spatial correlation often extends to the matching of fine-scale geometric features (e.g., presynaptic climbing fibers and postsynaptic Purkinje cell dendrites in the cerebellum). This obligatory correlation between pre- and post-synaptic elements, when framed in the context of circuit connectivity, provides a biologically relevant coordinate for morphological analyses.

The mesoscale correlation between pre- and post-synaptic elements is particularly identifiable and compelling for AAC and PyNs. Within the laminar architecture of the neocortex, different types of PyNs that project to distinct cortical and subcortical targets are organized, to the first approximation, into different layers, and different sources of cortical and subcortical inputs are routed through laminar streams (K. D. Harris & Shepherd, 2015) (Figure A1.3D). Importantly, the obligatory relationship between AAC axon terminals and PyN AIS represents a rare case where AAC axon distribution alone indicates connectivity to specific types of postsynaptic targets. Together, these provide an inherent spatial coordinate system to register AAC position and morphology in the framework of cortical input and output streams (Figure A1.3D). As the laminar arrangement of AAC dendrites recruit different input streams and the laminar stratification of axons mediate their output to separate PyN ensembles, we designed an AAC clustering analysis that emphasized the laminar density distribution of AAC dendritic and axonal arbors (Figure A1.6). We excluded L3 and L4 AACs

(Figure A1S2A) from this analysis as there were few such examples (less than 4) in our current dataset.

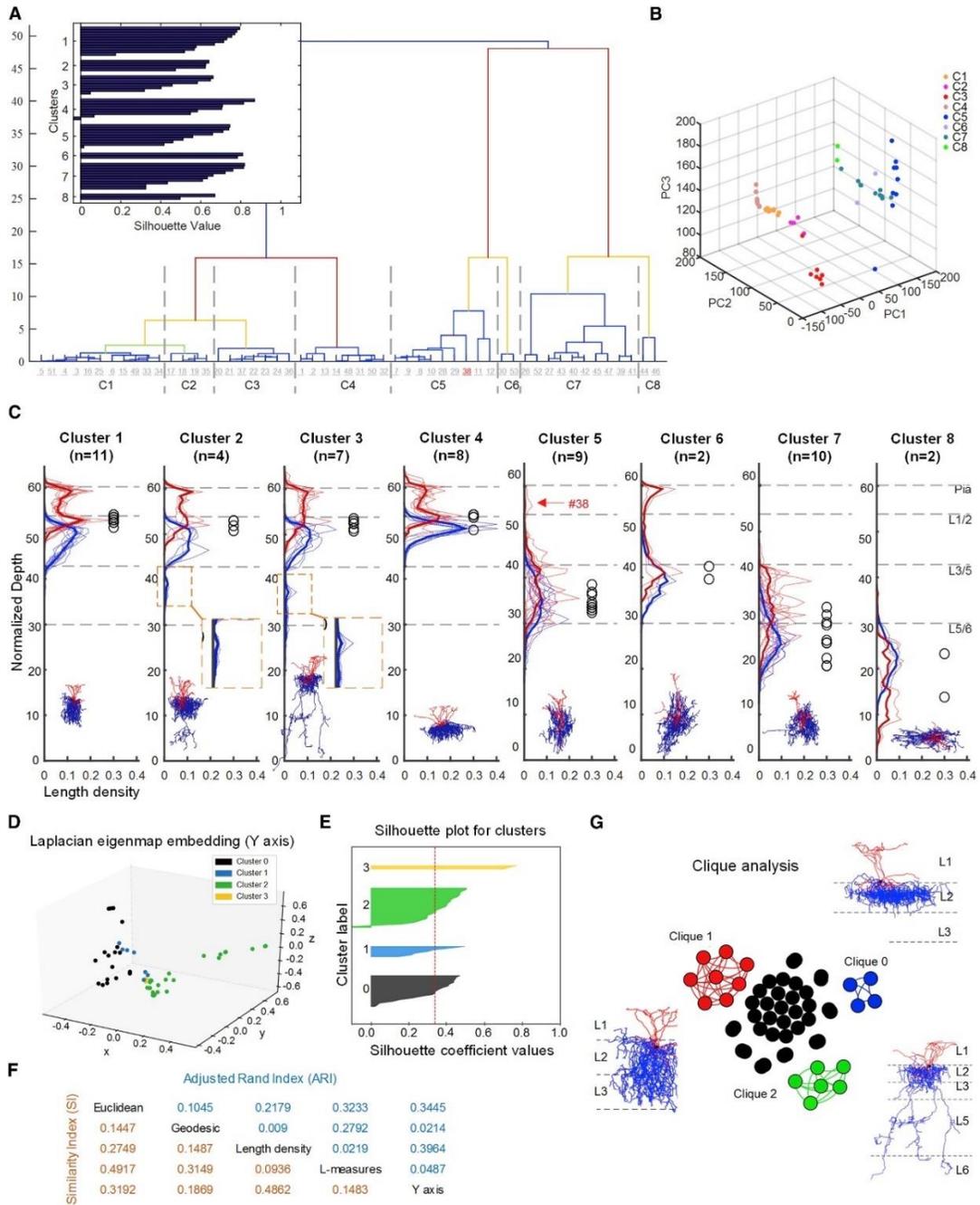
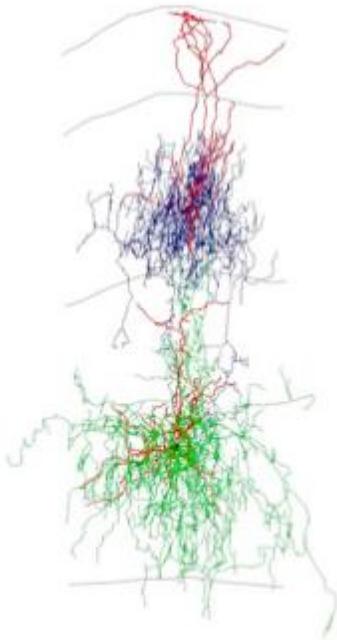


Figure A1. 6. Hierarchical Clustering of AACs Based on Cortical Laminar Density Distribution of Axons and Dendrites. (A) Dendrogram of hierarchically clustered AACs (n = 53). KL divergences (Kullback-Leibler divergence) of normalized arbor distribution functions along cortical depth were taken as the distance metric, and furthest distance was taken as the linkage rule. See more details in [STAR Methods](#) and [Figure A1S6](#) for the normalization procedures. Dashed lines correspond to the cutoff linkages of the identified eight cell clusters. Inset: silhouette analysis of the eight AAC clusters. (B) 3D scattering plots of the eight AAC clusters from (A) based on three principal components. (C) Axon (blue) and dendrite (red) length density distribution profiles of the eight AAC clusters. Dashed lines: cortical layer boundaries. Black circles: soma body positions. Bold lines: average of all the neurons in each cluster. Note that cell #38 in cluster 5 has apical dendrites (arrow) reading L1, a defining feature of cluster 6, but its lack of L3 axon branches (as it is located in SSC with a prominent L4) likely assigned it to cluster 5. (D–G) Clique analysis for the identification of robust AAC clusters. Clique analysis was conducted based on hierarchical clustering with five different metrics on AAC axons: three persistent-homology-based metrics, using three different ways of measuring distance from the soma, as scalar descriptor functions defined on the neuronal processes: Euclidean, geodesic, and depth from cortical surface (“y axis”), and the length density and L-measure metrics (Scorcioni et al., 2008b) defined in the text (D and E). Laplacian eigenmap embedding of hierarchical clustering for the y axis-based metric (D) and other descriptors ([Figure A1S7A–A1S7E](#)). The selection of “K” was based on silhouette analysis. Silhouette plot for K = 4 with y axis as the input metric; thickness denotes sizes of clusters; red dotted line denotes average silhouette score; larger score means better clustering (E). The relations between the five metrics were quantified by similar index (SI) and adjusted Rand index (F). Three robust AAC clusters were identified based on the clique analysis (G). The full listing of the three cliques are shown in [Figure A1S7F](#).

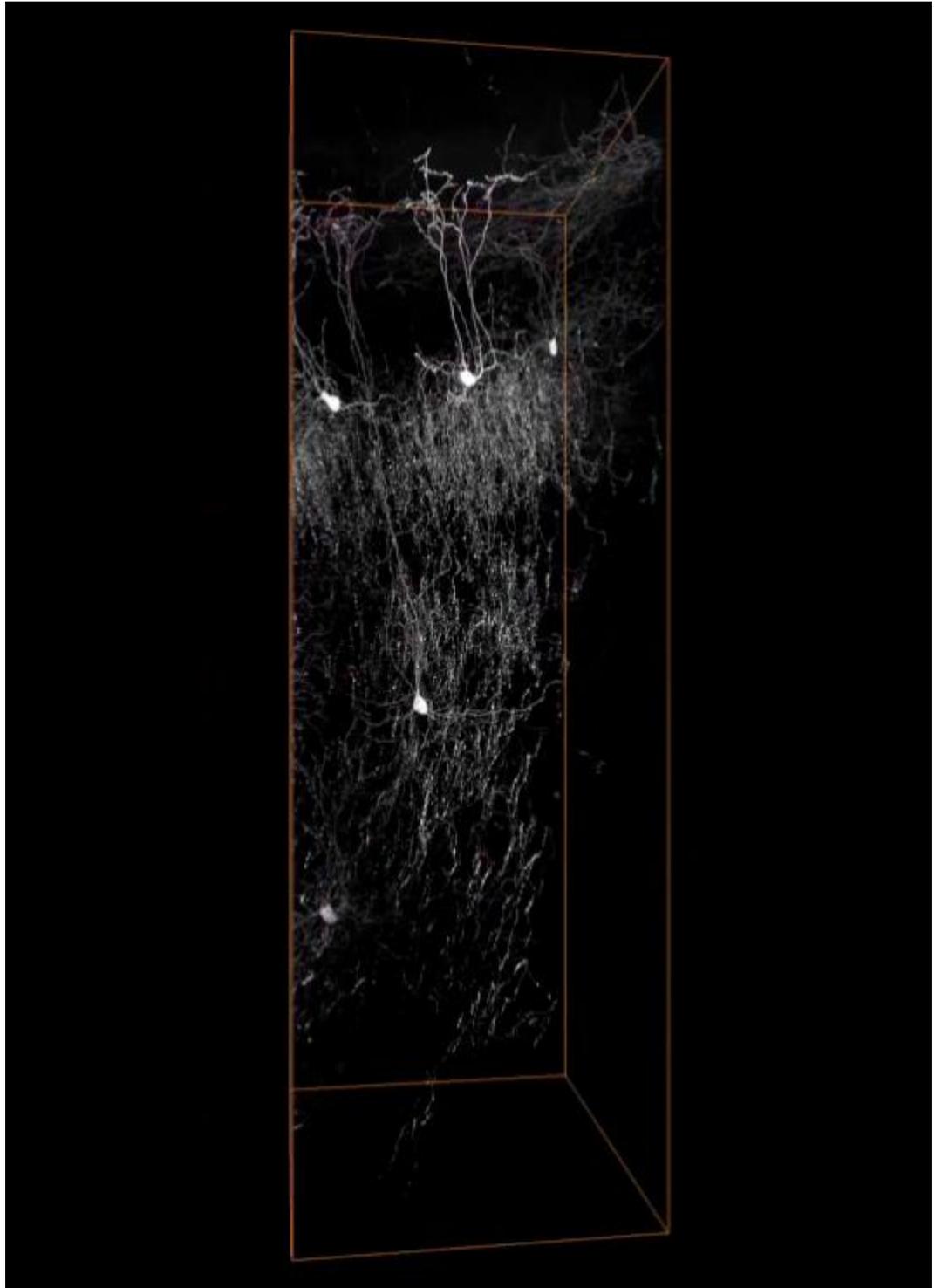
Based on brain cytoarchitecture information of dfMOST datasets, we normalized AAC dendrite and axon density distribution to a standardized cortex template ([Figure A1S6](#)). Hierarchical clustering based on cortical laminar density distribution of axons and dendrites revealed eight AAC clusters grouped according to the laminar distribution of their cell body position and dendritic and axonal arborization ([Figures A1.6A–A1.6C](#)). The four L2 ChC clusters correspond to intra- (cluster 4), cross- (cluster 1), and trans- (cluster 2 and cluster 3) ChC subtypes. The axon arbors of cluster 3 extend both L5 and L6 branches, but more dominantly innervate L5 ([Figure A1S2B](#)). Cluster 5 AACs resided in L5; their axons and dendrites were largely restricted within L5. Cluster 6 AACs resided in L5a; their axon arbors elaborated mostly within L5, but their dendrites

extended to L1. Cluster 7 AACs resided at L5 and L6 border (i.e., L6a), their dendrites were restricted in L5 and L6a, and their axons arborized mainly in L6. Cluster 8 consisted of L6 AACs with intralaminar dendrite and axon arbors. These different AAC clusters likely receive different inputs and control different subsets of PyNs, and thus are distinguished by their circuit connectivity patterns. Importantly, both the dendritic and axonal arbors of AACs conform to specific cortical layers rather than extending diffusely, presumably to recruit inputs or innervate targets in those layers, respectively. This suggests that the morphological variation of AACs might be more unitary than continuous. This was particularly apparent for two broad groups of L5 AACs, one extended short, L5-restricted apical dendrite and the other extended long, L1-reaching apical dendrites ([Figures A1.3](#) and [A1S5A–A1S5C](#)). We noted that this clustering method was not perfect as it assigned cell 38 to cluster 5, even though cell 38 extended apical dendrite to L1, as those characteristic to cluster 6 ([Figures A1.6](#) and [A1S5A](#)). In addition to these eight clusters ([Figure A1.6](#)), we detected three L3 AACs (two in SSC, one in MC) with translaminal axon arbors and apical dendrite reaching L1 ([Figures A1S5D](#) and [S5F](#); [Video S4](#)), one L4 AACs in SC ([Video S6](#)), and one inverted L6 AAC in mPFC ([Figures A1S5E](#) and [A1S5G](#) show the projection image of L6 AAC).



[Download video](#)

Supplementary Video A1.4. Reconstructions of Nearby L3 and L5-Intra AACs in MC, Related to Figure A1.3.



[Download video](#)

Supplementary Video A1.6. Nearby L4 and L6a (L5/L6 border) AAC in SSC, Related to Figure A1.3.

Hierarchical clustering based on the laminar distribution of axon density alone has a potential shortcoming: it may over-cluster or mis-cluster two identical density profiles appearing at different layer depths. Furthermore, low-dimensional projections do not always show well-separated clusters and may need other indirect evidence about clustering in the high-dimensional space such as silhouette plots. We therefore carried out a robust comparative analysis of morphological types using additional geometrical and topological characteristics of the neurons. For analyzing topological characteristics, we used a recently developed framework employing persistent homology (Y. Li, Wang, et al., 2017a) to derive a metric in the space of neuronal shapes (see [STAR Methods](#) and references). Briefly, this framework employs a descriptor function defined on the neurons, and a topological summary independent of neuronal location and orientation is derived from the descriptor function. We utilized three descriptor functions based on three different ways of measuring distances from the soma (Euclidean, geodesic, and cortical depth). In addition, we also used a community-standard metric (Scorcioni et al., 2008b), employed on <http://neuromorpho.org/>.

We performed hierarchical clustering employing each of these metrics, varying the number of clusters. By examining the overlap between the resulting clusters (adjusted Rand index [ARI] and similarity index [SI]; [Figures A1.6D and A1.6E](#)), we concluded that the metrics carry independent information about neuronal shape. We hypothesized

that if a pair of neurons appears in the same cluster across all the metrics, this provides robust evidence that those neurons belong to the same morphological type. We thus proceeded by defining a graph in which each neuron is a node, and two nodes are connected if and only if they appear in the same cluster across all five metrics considered. This procedure produced a set of disconnected cliques (fully connected clusters). The three largest cliques corresponded to three robustly identified AAC cell types that are also visible in the hierarchical clustering using only the laminar density of the axons: the intra-, cross-, and trans-L2 AAC subtypes ([Figures A1.6G and A1S7](#)). Currently, a number of AACs could not be grouped into cliques, likely due to less than enough sample size. We hypothesize that, with larger datasets, we will obtain similar robust cliques corresponding to other AAC subtypes for which evidence is provided by the hierarchical clustering shown in [Figures A1.6D–A1.6G](#).

AAC Subtypes Can Be Revealed by Axon Terminal Geometry That Correlates with That of Postsynaptic AIS

In addition to the laminar stratification of axon arbors, AAC axon terminals in different cortical layers manifested different geometric characteristics such as orientation, tortuosity, path distance, and branch order ([Figures A1.5E–A1.5H and A1.7](#)). As strings of AAC presynaptic terminals (i.e., “cartridges”) mostly align with the AIS of postsynaptic PyNs, we hypothesized that certain geometric features of AAC terminals reflect and correlate with those of the AIS. For example, the orientations of AIS in supragranular layers of mPFC were largely vertically aligned, but deviated substantially from this columnar orientation in infragranular layers (especially in L6; [Figures A1.7A–](#)

A1.7C and A1S2B). Consistent with this postsynaptic feature, AAC axon terminals in supragranular layers were also organized in predominantly vertical and parallel orientations, each largely straight and decorated with strings of presynaptic boutons (e.g., cartridges), which together earned them the name “chandelier cell.” In infragranular layers, on the other hand, the orientation of AAC terminals varied significantly with increased tortuosity that correlated with local PyN AISs (Figures A1.7D–A1.7I). Interestingly, analysis of several geometrical features of AAC terminals properly grouped AACs according to areas, laminar positions, and L2 subtypes (Figure A1.7J). In particular, several pairwise combinations of features classified AACs according to their areal, laminar locations, and even the three subtypes within layer 2 (Figures A1.7K–A1.7O). It is notable that AAC subtypes identified by axon local geometry are consistent with those identified by analyzing the laminar distribution of dendritic and axonal arbors (Figure A1.6), both rooted in their connectivity to PyNs. Together, these results suggest that a connectivity-based framework of morphological analysis is informative in resolving the granularity of AAC subtypes.

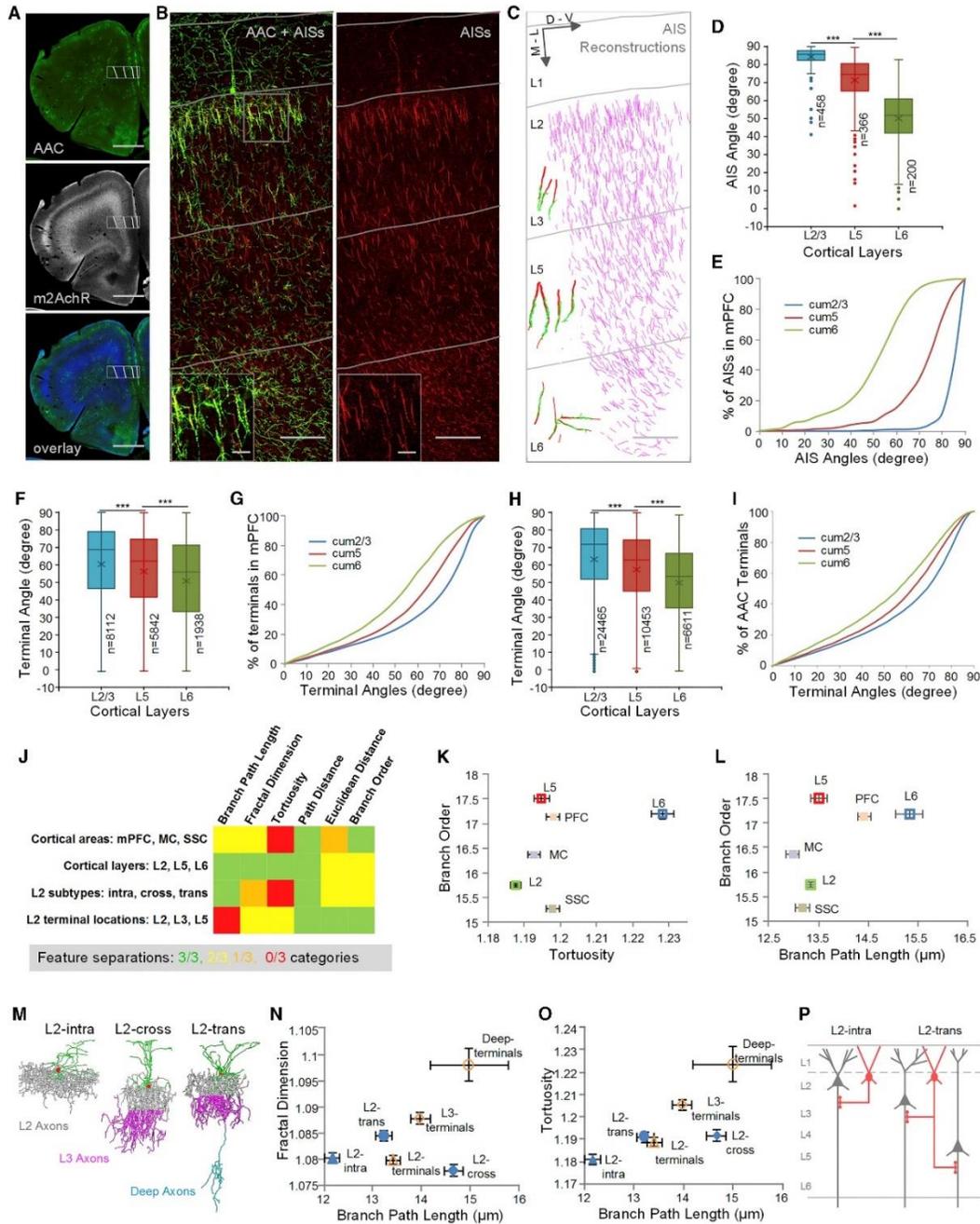


Figure A1.7. AAC Subtypes Revealed by Axon Terminal Characteristics That Correlate with AIS.

(A) Distributions of AACs in mPFC (50 µm thick). AACs were labeled by the crossing of Nkx2.1-CreER mouse and Ai14 (LSL-tdTomato) mouse with low dose of TM induction at E18.5. Top: AACs (green) shown by the immunostaining of tdTomato. Center: cortical layers shown by the immunostaining of m2AChR. Bottom: color merged. Scale bar: 1,000 µm. (B) AIS distributions in PrL (prelimbic cortex). Images were captured from the box in (A). Left: overlay image of AAC (green) and AIS (red). Right: immunostaining of AISs with Ankyrin-G. Insets: enlarged images. Gray lines indicate layer boundaries. Scale bars: 100 µm (low-mag) and 20 µm (insets).

(C) AIS reconstructions (purple). Insets: representative reconstructions of presynaptic AAC cartridge and postsynaptic PyN AIS pairs in L2/3, L5, and L6. Green: cartridges. Red: AISs. Scale bar: 100 μm . (D and E) Distribution differences of AIS angles among cortical layers in mPFC (**p < 0.0001, 95% confidence level, Kruskal-Wallis test followed by Dunn's multiple-comparisons test). Plots indicate median (full horizontal bar), mean (\times), quartiles, and range. AIS data (D) and cumulative plots (E) are from the reconstructions in (C). (F and G) The corresponding distribution differences of AAC axon terminal angles in mPFC (**p < 0.0001, 95% confidence level, Kruskal-Wallis test followed by Dunn's multiple-comparisons test). Plots indicate median (full horizontal bar), mean (\times), quartiles, and range. Axon terminal data (F) and cumulative plots (G) are extracted from all our AAC reconstructions in mPFC ([Figure A1S2A](#)). (H and I) Averaged distribution differences of AAC axon terminal angles across mPFC, MC, and SSC (**p < 0.0001, 95% confidence level, Kruskal-Wallis test followed by Dunn's multiple-comparisons test). Plots indicate median (full horizontal bar), mean (\times), quartiles, and range. Axon terminal data (H) and cumulative plots (I) are extracted from all our AAC reconstructions ([Figure A1S2A](#)). (J) Summary of axon terminal geometric features that separate AAC categories (cortical areas and cortical layers refer to somatic location). Green: statistically significant differences between all three pairs compared. Yellow: statistically significant differences between two of three pairs compared. Orange: statistically significant differences between one of three pairs compared. Red: no statistically significant differences. (K and L) Areal and laminar categories of AACs separated by axon terminal geometric parameters. Parameter pairs of branch order and (K) branch tortuosity or (L) branch path length are shown. Data are mean \pm SEM. (M) Reconstructions of representative L2 AAC subtypes (L2-intra, L2-cross, and L2-trans). (N and O) Axon terminal geometric parameters separate L2 AAC subtypes. Parameter pairs of branch path length and (N) branch fractal dimension or (O) branch tortuosity are shown. Data are mean \pm SEM. (P) Schematic of inferred L2 AAC connectivity with PyNs.

Discussion

As individual neurons are the basic building blocks of the nervous system, single-neuron analysis is essential to reveal the true degree of cell diversity and principles of circuit organization. Morphology is an intuitive depiction of neuron types that reflects their input-output connectivity; thus, the visualization and quantification of complete single-neuron shapes are necessary to identify and classify neuron types and deduce their anatomic relationships. However, the vast diversity, large spatial span, and vexing variations of mammalian neurons present unique challenges in cell labeling, imaging, and analysis. Recent advances in light microscopy begin to overcome the technical hurdle of submicron-resolution imaging of the entire mouse brain using either wide-field structured

light illumination microscopy (i.e., MOST and fMOST) (Gong et al., 2013; A. Li et al., 2010) or fast-scanning two-photon microscopy (Economo et al., 2016). In particular, the dual-channel dfMOST approach allows fast and simultaneous acquisition of both neural structures and their whole-brain spatial reference at cellular resolution (Gong et al., 2016).

Another key requirement for reconstructing single neurons using light microscopy is sparse and robust labeling, and systematic labeling across neuronal populations is necessary to achieve comprehensive discovery of neuron types. Conventional transgenic approach lacks specificity and sparseness. Although viral vectors can achieve sparse labeling of distal axons (Economo et al., 2016), their limitations include (1) dense labeling of local collaterals that are difficult to reconstruct, (2) lack of specificity to local interneurons, and (3) lack of orthogonal information (e.g., molecular markers) to further restrict labeling and help interpret morphological variations in cell type identification. Our combinatorial genetic strategy overcomes these limitations. We engage multiple cell features to target subpopulations defined by gene combinations, lineage, birth time, and anatomy (He et al., 2016). We further incorporate inducible and viral methods to achieve reliable single-cell labeling (Figure A1.1B), which should enable “saturation screening” of morphological types or subtypes within the subpopulation. Although here we have not reached saturation screening of cortical AACs, as L4 and inverted L6 AACs were detected only once in our dataset, the approach demonstrates unprecedented specificity and comprehensiveness to a rare cortical cell type. Iterations of this labeling scheme through systematic generation of mouse

driver lines (J. A. Harris et al., 2014) will enable comprehensive targeting of neuron types as has been demonstrated in genetic targeting in *Drosophila* (Aso et al., 2014; Jenett et al., 2012). Thus together with scRNA-seq, gSNA provides an orthogonal and scalable single-neuron analysis platform. Currently, the bottleneck of the gSNA is single-neuron reconstruction, which mostly relies on manual procedures. Future innovation in machine learning-based automatic reconstruction algorithms may increasingly overcome this limitation (Peng et al., 2015, 2017a).

The goals of single-neuron anatomy are to identify and catalog cell types and, ultimately, to inform cell function through inferring connectivity. With increasing throughput in single-neuron reconstruction, a pressing issue is how to extract biologically relevant information from morphology datasets. Traditional approaches deploy a large set of geometric and topological metrics (The Petilla Interneuron Nomenclature Group (PING), 2008) to quantify single-neuron morphology in isolation often without a proper spatial coordinate and circuit context; these analyses are mostly ineffective in parsing neurons into reliable and biologically informative groups. As morphology is a proxy to and serves the purpose of connectivity (Seung & Sümbül, 2014; Sümbül et al., 2014), we have adopted a connectivity-guided approach to morphological analysis. This approach is based on the premise that, although single-neuron shape by itself does not contain information about presynaptic sources and postsynaptic targets, such information can be extracted, to varying degrees, if neuron morphology can be registered and analyzed in an appropriate spatial coordinate that reflects local and/or global connectivity patterns. Indeed, the inherent polarity of dendrites and axons ensures that their distribution and

geometry reflect the input source and output targets in the corresponding spatial domains, or “anatomic parcels” (Ascoli & Wheeler, 2016). Although the precise identity of input and output elements cannot be inferred from spatial location alone, anatomic parcels based on decades of classic studies provide significant information to include and exclude pre- and post-synaptic elements and thus to infer possible as well as impossible connectivity. This analysis framework is likely to recognize seemingly “subtle” morphological variations (e.g., translaminal dendrite or axon branches of AACs), which yet have significant impact on connectivity and thus cell function. In this context, the dfMOST datasets, which allow automatic registration of single-neuron morphology into proper global and local coordinates at cellular resolution within the same brain (Gong et al., 2016), is key in analysis strategies to identify and distinguish cell types and to inform connectivity. In analyzing AAC morphology, for example, the precise cell distribution information of the dfMOST dataset is crucial to derive and normalize laminar coordinates in different cortical areas, which enabled areal and laminar comparisons and inferences of input-output connectivity patterns that distinguished AAC subtypes.

A recent study suggests that cardinal GABAergic neuron types are distinguished by their input-output synaptic communication patterns encoded in transcription profiles (Paul et al., 2017). Beyond cardinal types, finer division into subtypes may be necessary to represent and explain the intricacies of neural circuit organization (Zeng & Sanes, 2017), but there is no consensus and mechanistic basis on the granularity and boundary of neuronal subtypes. Our results on cortical AACs suggest that differences in input-output connectivity, which are reflected in cell morphology, are likely a major determinant of

neuronal subtypes. Differential gene expression in supragranular or infragranular AACs in the frontal cortex (Paul et al., 2017) is consistent with this interpretation. We therefore suggest that synaptic communication patterns may distinguish neuronal subtypes as well as cardinal types.

It is notable that AAC subtypes, when registered along cortical laminar coordinates, appear to manifest a degree of stereotypy and fine granularity that is similar to those of retinal bipolar cell subtypes registered upon the much finer coordinates of retinal sub-laminae (Shekhar et al., 2016). A true saturation anatomical analysis of cortical AACs will likely reveal additional subtypes. While the division of retinal bipolar subtypes is further supported by molecular, physiological, and functional evidence (Euler et al., 2014), the division of AACs subtypes based on anatomy needs to be substantiated by orthogonal datasets, such as their physiological connectivity (J. Lu et al., 2017) and gene expression profiles (Paul et al., 2017). On the other hand, our results suggest that high-resolution morphology dataset alone, when registered within proper spatial coordinates that reflect brain circuit organization, contain rich anatomical information on cell identity and connectivity, providing a structure basis to integrate orthogonal datasets. This analysis strategy should apply to projection neurons as dfMOST datasets contain brain-wide information on anatomical parcels that will inform the potential synaptic targets of long-range axon branches. Therefore, light microscopy-based high-throughput single-neuron anatomy will likely provide substantial information and insight on cell type diversity and mesoscale connectivity in the mammalian brain.

STAR★Methods

Key Resources Table

Table A1. 1. Key resource table.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal RFP	Rockland	Cat# 600-401-379; RRID: AB_828392
Mouse monoclonal Ankyrin-G	Neuromab	Cat# 75-146; RRID: AB_10673030
Rat monoclonal muscarinic Acetylcholine receptor m2 (m2AChR)	Millipore	Cat# Cat# AB5166; RRID: AB_91715
Alexa Fluor 488 goat anti-rat	Invitrogen	Cat#A-11006; RRID: AB_2534074
Alexa Fluor 594 goat anti-rabbit	Invitrogen	Cat# A-11037; RRID: AB_2534095
Alexa Fluor 647 goat anti-mouse IgG2a	Invitrogen	Cat#A-21241; RRID: AB_2535810
Recombinant DNA		

pAAV-EF1a-fDIO-TVA-GFP	this paper	N/A
pAAV-EF1a-FLEX-GT	Callaway Lab	Addgene plasmid # 26198
pAAV-EF1a-fDIO-YFP-WPRE	Deisseroth Lab	<u>Fenno et al., 2014</u>
Chemicals, Peptides, and Recombinant Proteins		
Lowicryl HM20 resin	electron microscopy sciences	Cat#RT-14340
Propidium iodide (PI)	Invitrogen	Cat#P1304MP
Experimental Models: Organisms/Strains		
Nkx2.1-CreER mouse	Jackson Laboratory	JAX: 014552
Rosa26-loxp-stop-loxp-flpo (LSL-Flp) mouse	Jackson Laboratory	JAX: 028584
Rosa26-lox-stop-lox-tdTomato (Ai14) mouse	Jackson Laboratory	JAX: 007905
Software and Algorithms		
Amira	FEI, Mérégnac Cedex, France	https://www.thermofisher.com/global/en/home/industrial/electron-microscopy/electron-microscopy-instruments-workflow-solutions/3d-visualization-analysis-software/amira-life-sciences-biomedical.html ; RRID:SCR_014305
NeuroLucida360	MBF Bioscience, Williston, VT	https://www.mbfioscience.com/neuroLucida360 ; RRID:SCR_001775
NLMorphologyConverter	Neuronland; Neuromorpho.org	http://neuronland.org/NLMorphologyConverter/NLMorphologyConverter.html ; RRID:SCR_001817

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Qingming Luo (qluo@mail.hust.edu.cn) or Z. Josh Huang (huangj@cshl.edu).

Experimental Model and Subject Details

Experimental Animals and Low Dose TM Induction

To achieve sparse and specific targeting of AACs across neocortical areas, we crossed *Nkx2.1-CreER* mice (The Jackson Laboratory stock 014552) with *Rosa26-loxP-stop-loxP-flpO* (*LSL-Flp*) mice (The Jackson Laboratory stock 028584). At postnatal day 0 (P0) or day 1 (P1), we intraperitoneally induced each pup with low dose of tamoxifen (TM, 0.25 mg per pup). Tamoxifen stock solution (5mg/ml in corn oil) were prepared beforehand. Sparsely targeted AACs would express FlpO constitutively (He et al., 2016).

For immunostaining experiments, we crossed *Nkx2.1-CreER* mice with *Rosa26-lox-stop-lox-tdTomato* (*Ai14*) mice (The Jackson Laboratory stock 007905). To ensure embryonic day 18.5 (E18.5) TM inductions, Swiss Webster or C57B6 females (Taconic) were housed with *Nkx2.1-CreER: Ai14* (ht/homo) males overnight and females were checked for vaginal plug by 9am the following morning. At E18.5, pregnant females were given oral gavage administration of TM (dose 3mg / 30 g of body weight) for sparse

labeling of AACs. AACs are labeled with tdTomato. Genetic hybrids of C57B6 and Swiss Webster animals were used in these experiments. All animal breeding and surgical experiments were approved by the Institutional Animals Care and Use Committee (IACUC) of Cold Spring Harbor Laboratory or the Institutional Animal Ethics Committee of Huazhong University of Science and Technology.

Method Details

Stereotaxic Virus Injection

Flp dependent pAAV-EF1a-fDIO-TVA-GFP (TVA: avian glycoprotein EnvA receptor) cassette was assembled and cloned using standard molecular cloning protocols with restriction enzymes from New England Biolabs. TVA-GFP (*pAAV-EF1a-FLEX-GT*) was a gift from Ed Callaway (Addgene plasmid # 26198). The cassette was subcloned into *pAAV-EF1a-fDIO-YFP-WPRE* (a gift from the Deisseroth laboratory, Stanford University) using NheI and AscI cloning sites (Fenno et al., 2014). All constructs were sequenced to ensure their fidelity and proper reversed orientation of the inserts, and packed into AAV8 viral vectors with titers ranging from 1.0×10^{12} to 2.4×10^{12} pfu from the UNC Vector Core (*Chapel Hill, North Carolina*).

For stereotaxic injection, post-weaned animals (3 to 4-week-old) were anesthetized by intraperitoneal injection with ketamine and xylazine (ketamine:100 mg/kg, xylazine: 10 mg/kg in saline), and then were fixed in a stereotaxic headframe (*Kopf Instruments Model 940 series*) for the identification of the coordinates of mPFC, MC and SSC areas based on the

Allen Mouse Brain Reference Atlas (<http://atlas.brain-map.org>). Each animal received bilateral injection in mPFC, MC and SSC areas (6 injection sites per mouse). At each site, we injected 100 nL virus with Nanoliter 2010 Injector (World Precision Instruments). And we let virus express more than 21 days for strong labeling. The membrane tagged labeling by TVA-GFP fusion significantly improved the labeling of fine axon terminal. The stereotaxic coordinates are: mPFC (A/P: 1.98 mm, M/L: \pm 0.5 mm; D/V: 1.5mm depth from pial surface), MC (A/P: 0.5 mm, M/L: \pm 1.5 mm; D/V: 0.5 mm) and SSC (A/P: -1.5 mm, M/L: \pm 3.0 mm; D/V: 0.5mm).

Immunostaining

Animals (P45-P60) were perfused with 4% PFA in PBS. The brains were removed and post-fixed overnight in the same fixative. Coronal brain slices were sectioned at 75 μ m thickness via vibratome. Sections were blocked with 10% normal goat serum in 0.5% Triton in PBS for an hour and then incubated overnight with primary antibodies diluted in blocking solution at room temperature. Primary antibodies used were: rabbit polyclonal RFP (1:1000, Rockland) for labeling AACs, mouse monoclonal Ankyrin-G (1:500, Neuromab) to label pyramidal axon initial segments (AIS), and rat monoclonal muscarinic Acetylcholine receptor m2 (m2AChR) (1:500, Millipore Sigma) to discriminate L3/5 and L5/6 boundaries in mPFC. Sections were subsequently washed and incubated with the appropriate fluorescently-conjugated secondary antibodies diluted in the same buffer for 3 hours at room temperature. Secondary antibodies used were: Alexa Fluor 488 goat anti-rat (1:500, Invitrogen), Alexa Fluor 594 goat anti-rabbit (1:500, Invitrogen), and Alexa Fluor 647 goat anti-mouse IgG2a (1:500, Invitrogen).

Perfusion and Whole-Brain Resin Embedding (Gang et al., 2017)

Mice were deep anesthetized by overdose of ketamine and xylazine, and then intracardially perfused with 0.01M PBS (Sigma-Aldrich Inc., St Louis, MO, USA) and 4% paraformaldehyde (PFA, Sigma-Aldrich Inc., St Louis, MO, USA). After brain dissection and about 18 hours of post-fixing in 4% PFA, brain samples were rinsed in 0.01M PBS overnight. Then samples were dehydrated in graded series of ethanol (with distilled water): 50% ethanol (2h, 3 times), 75% ethanol (2h, 1 time), 100% ethanol (2h, 2 times). After dehydration, we replaced ethanol with graded series of xylene (with pure ethanol): 50% xylene (2h, 2 times) and 100% xylene (2h the first time, and then overnight). We then infiltrated samples in graded series of Lowicryl HM20 resin (in xylene): 50% HM20 (2h), 75% HM20 (2h), 100% HM20 (2h, 2 times), 100% HM20 (2 days). After resin infiltration, samples were heat-polymerized at 50°C for 8 hours in a vacuum oven. All dehydration and infiltration procedures were treated at 4°C. All solutions were prepared in weight.

Note: During wide-field based dfMOST imaging, autofluorescence produced by lipofuscin in the resin-embedded brain tissue often interfered with image contrast. Swiss Webster mice, especially after 2 months of age, usually express more lipofuscin compared with C57/BL6 mice. To reduce the effect of lipofuscin, all animals in this study were sacrificed around P51-P54.

Whole-Brain Dual-Color fMOST (dfMOST) Imaging

Plastic embedded brain samples were mounted on a metal base and then installed under a dual-color fluorescence micro-optical sectioning tomography (dfMOST) system

for whole-brain imaging. The dfMOST system is a wide-field block-face imaging system. A blue laser (488nm) was used as the excitation light source with two separate TDI-CCD cameras for signal detection. This system runs in a stripe scanning mode (x axis) and combines with an afterward image montage to realize the centimeter-scale coronal data acquisition (T. Yang et al., 2015). A precision motorized XYZ stage is used to conduct imaging scanning, areal expansion and ultra-thin sectioning by a diamond knife (A. Li et al., 2010). The high throughput and high resolution imaging of fluorescent protein labeled samples is realized with chemical sectioning (X.W. and T.Y., unpublished data). Following each scanning of one coronal plane (X-Y axes), the sample was sectioned to remove the top layer (Z axis), and then imaged again. The imaging-sectioning cycles were performed automatically with 1.0 μm Z steps until whole brain was imaged. The resin-embedded GFP fluorescence were well preserved through chemical reactivation (Xiong et al., 2014) provided by adding Na_2CO_3 in the imaging water bath (0.05 M, PH = 11.4).

We used a 60X water immersion objective (NA 1.0) for imaging, which provided the system with submicron resolution at $0.2 \times 0.2 \times 1 \mu\text{m}$ voxel sampling rate for each whole-brain dataset. High resolution and high density sampling rate greatly facilitated our cell reconstruction procedures and are especially necessary for reconstructing dense neural arborizations and fine structures (such as axon boutons and spines).

The red channel was used to capture the whole brain cytoarchitecture which was counterstained by propidium iodide (PI) (Gong et al., 2016). PI dyes were dissolved in the imaging water bath, thus stained the exposed DNAs and RNAs on the tissue surface.

The staining occurred within thus was in real time. The 488 nm laser was strong enough for PI excitation. The counterstained cytoarchitecture provided a self-registered Nissl like brain atlas for the GFP channel and was used for the identification of cortical areas and layers. Furthermore, the image contrast in the PI channel was sufficient for identification and reconstruction of PyN main dendrites ([Figures A1.3A and A1.3B](#)). Weakly stained or unstained tubular cellular objectives, such as blood vessels and pyramidal main dendrites, can be seen in good contrast in PI channel.

Quantification and Statistical Analysis

Single Cell Reconstruction and Layer Boundaries Discrimination

To reconstruct sparsely labeled single ChCs from the whole-brain image datasets (~8 TB), we transformed TIFF format raw images series to LDA type (Y. Li, Gong, et al., 2017). We then used Amira software (*v 5.2.2, FEI, Mérignac Cedex, France*) to load the LDA data and identify cells for initial reconstruction. We only chose cells with highly characteristic axon terminal cartridges which were true ChCs (~30% GFP-labeled neurons were not the ChC type). The areal and laminar location of selection of cells were identified based on the cytoarchitecture provided by PI staining according to Allen Mouse Brain Reference Atlas. All initial reconstructed cells were saved in SWC format.

The arborization of a complete single ChC was extremely dense: the average length of AAC axons was ~2.1cm, the average number of AAC axon branches was ~1369, and the average axon branch order was ~31. Only manual procedure was feasible to reconstruct cells with such arbor complexity. Each AAC took up to one week to complete by one person.

To ensure all AAC reconstructions were correct and complete, we carried out revisions on each initial reconstruction in NeuroLucida360 software (*NeuroLucida*, MBF Bioscience, Williston, VT). Since NeuroLucida 360 was not compatible with the reading of SWC format files and couldn't hold TB-size image datasets, we transformed all the SWC files to NeuroLucida ASC format using the Neuronland software (Neuromorpho.org), and we cropped smaller image stacks (GB-size) of GFP and PI channels from the whole brain datasets. The Cropping areas were based on the coordinates calculated from the initial SWC reconstructions.

Cortical layer boundaries were reconstructed in the co-registered PI channel in NeuroLucida 360. Laminar positions were discriminated based on cell body distributions according to the online version of Allen Mouse Brain Atlas (<http://portal.brain-map.org/>). 5 μ m max intensity projections of PI images were used to better show the cell body distributions. Since only PFC area has relative clear L2, L3 boundaries, we did not draw L2/L3 borders for all the cells. In our analysis, the axons are separated to layer 2 and layer 3 by defining the upper half of L2/3 as layer 2, and the lower half as layer 3. L3/5 boundary was identified by the existence of sparser cell distributions and larger pyramidal cell bodies in L5. L5/6 boundary was determined based on the missing of large cell bodies and the appearance of denser and horizontal oriented cell bodies in L6.

Adjusting the Orientation AACs to the Vertical Axis of Local Cortical Column

To identify the local vertical axis of cortical depth, we randomly reconstructed a few pyramidal main dendrites near the reconstructed AAC cell body in PI channel. We

took the main direction of PyN apical dendrites near the AAC cell body as the cortical column vertical axis. We first randomly reconstructed a few pyramidal apical dendrites. We then centered all the traced pyramidal dendrites, and identified their main orientation based on the vectors calculated with pyramidal dendritic vectorization by Principal Component Analysis (PCA). Using this orientation as the proxy of cortical vertical axis, we re-orientated each AAC reconstruction using the MATLAB software (Figure A1.3B).

Length Density Analysis

Length density analysis of AAC morphology

Length density analysis of dendrites and axons were performed using custom MATLAB routines (Yamawaki et al., 2014). Briefly, for each orientation-readjusted AAC, we set the soma center as origin of coordinate. The neuronal arbors were divided into $15\ \mu\text{m} \times 15\ \mu\text{m}$ grid space in the xy plane, and the arbor length in each grid were summed covering the whole z direction. The distribution pattern in coronal plane (i.e., xy plane) was represented in heatmap. Length density profile along the cortical depth direction (i.e., y axis) were plotted to quantify the laminar distribution pattern by integrating fiber length along x direction from heat-map. Similarly, length density profiles along x axis (middle-lateral) and z axis (anterior-posterior) were plotted to quantify the horizontal distribution patterns (Figures A1S4C–A1S4F). To make easy comparison, we normalized profile by dividing the fiber total length of the cell (length ratio). Layer boundaries were also plotted in the length density figures (dashed lines). Their positions in the y axis was the average coordinates of all the contouring points covering the neuron arbor extent in x direction (Figure A1S6).

Normalized Length density distribution on a standard cortex template

For comparative analysis among AACs from different brain areas and samples, we normalized the laminar distribution of AAC axonal and dendritic arbors to a standard cortex template (y axis only). In the neocortex, only SSC has L4 compared with mPFC and MC, and the L6-WM (white matter) border in mPFC is usually not discernable in the coronal plane. And the thickness of the same layer in different cortical areas and even subareas can be different. To address these issues, we performed normalization based on the thickness of each layer, rather than using the distance from pia to WM. The number of laminar arbitrary units (AUs) been used for subdividing each layer was decided based on the average thickness of each layer from all cells (L1: 100.04 μm , L2/3: 180.52 μm , L5: 215.33 μm , $n = 53$). Here we kept the dividing size to be around 15 μm to match with the unnormalized length density analysis. Thus, the numbers of laminar AUs for dividing L1, L2/3 and L5 are 7 AUs, 12 AUs and 14 AUs respectively. According to these parameters, as shown in Figure A1S6, axon arbors were subdivided with different intervals for L1, L2/3 and L5. For the arbors above L1 and below L5 (L6), we used the dividing intervals of L1 and L5 respectively. Since the axons of most AACs did not innervate L4 (except one L4 AAC), we removed L4 length density data for all the AACs in SSC. Based on this method, we could get normalized length density distribution curves of dendrites and axons for each cell from all the three cortical areas.

Cluster Analysis

Based on normalized distribution of neural arbor length density along the y axis (cortical depth), 53 AACs were hierarchically clustered using a weighted KL divergence

(KLD_w , symmetrized (D. Johnson & Sinanovic, 2001)) as the distance metric and the furthest distance as the linkage rule. A weight coefficient λ was defined as the ratio of average axon length to total axon and dendrite length across all neurons. The KLD_w matrix was calculated by multiplying the axon distribution by the length based weight λ and the dendrite distribution by $(1 - \lambda)$. That is $KLD_w = \lambda KLD_{axon} + (1 - \lambda) KLD_{dendrite}$. Based on the clustering dendrogram of KL divergence, 53 AACs can be grouped to different clusters. Corresponding silhouette analysis was done based on the cutoff linkages used in the clustering

Clique Analysis

To robustly classify the AACs we did a comparative clustering study across five different metrics, to find neuronal groups that clustered together irrespective of metric utilized.

Three of the metrics were derived from topological considerations described in (Y. Li, Gong, et al., 2017; Y. Li, Wang, et al., 2017a). This methodology starts by defining a “descriptor function,” which is a scalar valued function defined on the axons and dendrites of each neuron. The procedure then computes a topological signature known as the persistence diagram for each neuron based on the descriptor function. Finally, the distance between two neurons is defined by computing a suitable metric between the persistence diagrams. The persistence diagrams are by definition invariant to rigid translations and rotations, and may have further invariances. Three of the metrics were defined by using three different descriptor functions, in each case defined as a suitable distance from the soma to the point on the neuron. The three distance functions

used were Euclidean distance, Geodesic distance along the neuron, and distance along the normal to the cortical sheet (we denote this the “y-axis” for brevity).

In addition, we used a metric defined by taking KL distance between the histograms created by projecting the neuronal processes onto the y axis (“length density”), and finally a community-standard metric, the L-measure (Scorcioni et al., 2008b), used on neuromorpho.org.

How related are these metrics? To answer this question, we performed hierarchical clustering using each of the metrics, fixing the total number of clusters to be K . In general, the different metrics produced different sets of clusters. We compared the sets of clusters across two metrics, using the Adjusted Rand Index (ARI), and the Similarity Index (Bohland et al., 2009) (SI) defined in Bohland et al. (<https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0007200>) to compare different parcellations of brain atlases. In each case, these indices lie between 0 and 1, with 1 corresponding to perfect correspondence between two sets of clusters. We found (Figure A1.6F) that both indices were generally closer to 0 than to 1, indicating that these metrics measured independent geometrical/topological characteristics of the neurons. Thus, if neurons were grouped together by all five metrics, we would gain confidence that they were robustly classified into these clusters.

To perform this robust classification, we used the following method: (i) first, we carried out hierarchical clustering using each of the metrics, with a fixed number K of clusters. (ii) We then defined an undirected graph G with each node corresponding to a neuron. The edge between two neurons is either 1 or 0 based on whether the neurons

clustered together or not as described below. (iii) We then looked for disjoint cliques (in a clique, each node is connected to every other node in the clique; intuitively, a clique constitutes a set of very similar neurons). These disjoint cliques were our robust clusters.

Let the number of metrics be M ($= 5$ in our case). We introduced a parameter N that controlled the edge weights of the graph G as follows: If two neurons belonged to the same cluster for at least N of the M metrics, then we gave that edge a weight 1, otherwise we gave it a weight 0. Thus, the graph G was a function of two parameters K, N . We then looked for maximal cliques in $G(K, N)$. For $N < M$, the maximal cliques in G were not in general disjoint, however for $N = M$ the cliques can be shown to be disjoint. Consider the relation between two neurons given by an edge in $G(K, M)$. This relation is transitive: if two pairs of neurons (N_1, N_2) and (N_2, N_3) are connected, then (N_1, N_3) must also belong to the same cluster (of which N_2 is a member). This transitivity guarantees the disjointness of the maximal cliques: If two cliques share a vertex, then the two cliques must be identical. Thus, we considered only $G(K, M)$ and found the disjoint maximal cliques. In our case $M = 5$. We selected K by examining the average silhouette scores of the clusters versus K (https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html). Finally, performing clique analysis on $G(K = 4, N = M = 5)$, we found 3 cliques with size greater than 2 (with sizes 4, 6 and 8 respectively; [Figures A1.6G and A1S7](#)). These cliques were the output of our robust clustering analysis, and exemplars from each clique are [shown in Figure A1.6G](#).

Calculation of Single Neuron Anatomical Features

Basic neuron morphological features of neuronal dendrites and axons were calculated using Neurolucida software. The position of cell body (area, layer) were practically identified based on the PI channel cytoarchitecture information. The distance of soma to L1/2 border were calculated based on the readjusted neuron orientation (Figure A1.3B) with MATLAB software. Area: brain areas that cell body stays. L1 thickness: the thickness of L1. Layer: laminar position of cell body. Soma to L1/2 border: the distance of soma center to L1/2 in micron. Soma radius: average of the distance of reconstructed contour points to the center. Dendrite Qty: the number of dendrite trees that grows out from soma. Dendrite nodes: the number of branching points of all dendrites. Dendrite ends: the number of terminal tips of dendrites. Dendrite full length: the full length of all dendrite fibers. Dendrite mean length: the mean length of all dendrite trees. Dendrite max_branch_order: max branch order of all dendrite trees. For branch order calculation, here we use centrifugal ordering, which is the basic scheme to assign branch order to a tree. Dendrite branch Qty: the number of total dendrite branches. Dendrite branch mean length: average length of all branches. Dendrite mean tortuosity: average of the tortuosity values of all dendrite branches, tortuosity is defined as [Distance along process] / [Straight line distance]. Dendrite tortuosity SD: standard deviation of the tortuosity values of all dendrite branches. Dendrite convex hull: the volume of a convex polygon by connecting the tips of the distal dendrites or axon. Axon morphological parameters are similarly defined as dendrite.

Sholl Analysis on Dendrites and Axons

We did sholl analysis on AAC dendrites and axon fiber distributions at 3D. Fiber length was used here for analysis. Sholl analysis generates a set of nested concentric spheres centered at the cell body. The smallest sphere has a radius of 5 um considering the soma radius. The spheres increase in size by a constant change in 15um, which is defined as the radius in [Figures A1.4G](#) and [A1.5C](#). According to this spheres, many shells are constructed. Shell is the volume contained out to the given radius, but does not include the volume of any smaller shells. The fiber length in each shell was summarized. To compare different cell, we normalized fiber length as [fiber length in a shell] / [the full length for a given fiber type].

Data and Software Availability

Code Availability

All custom codes used in this study are available from the corresponding author upon reasonable request.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China ([61721092](#) and [81827901](#)) and Director Fund of WNLO (Q.L., H.G., and S.Z.), 5R01MH094705-04 and CSHL

Robertson Neuroscience Fund (Z.J.H.), R01NS39600 and Burroughs Wellcome Fund Collaborative Research Travel Grant (G.A.A.), 1R01 EB022899-01 (Y.W. and P.M.), and a Crick-Clay Professorship (P.M.). J.T. and B.-S.W. were supported by NRSA pre- and post-doctoral fellowships, respectively.

Supplemental information

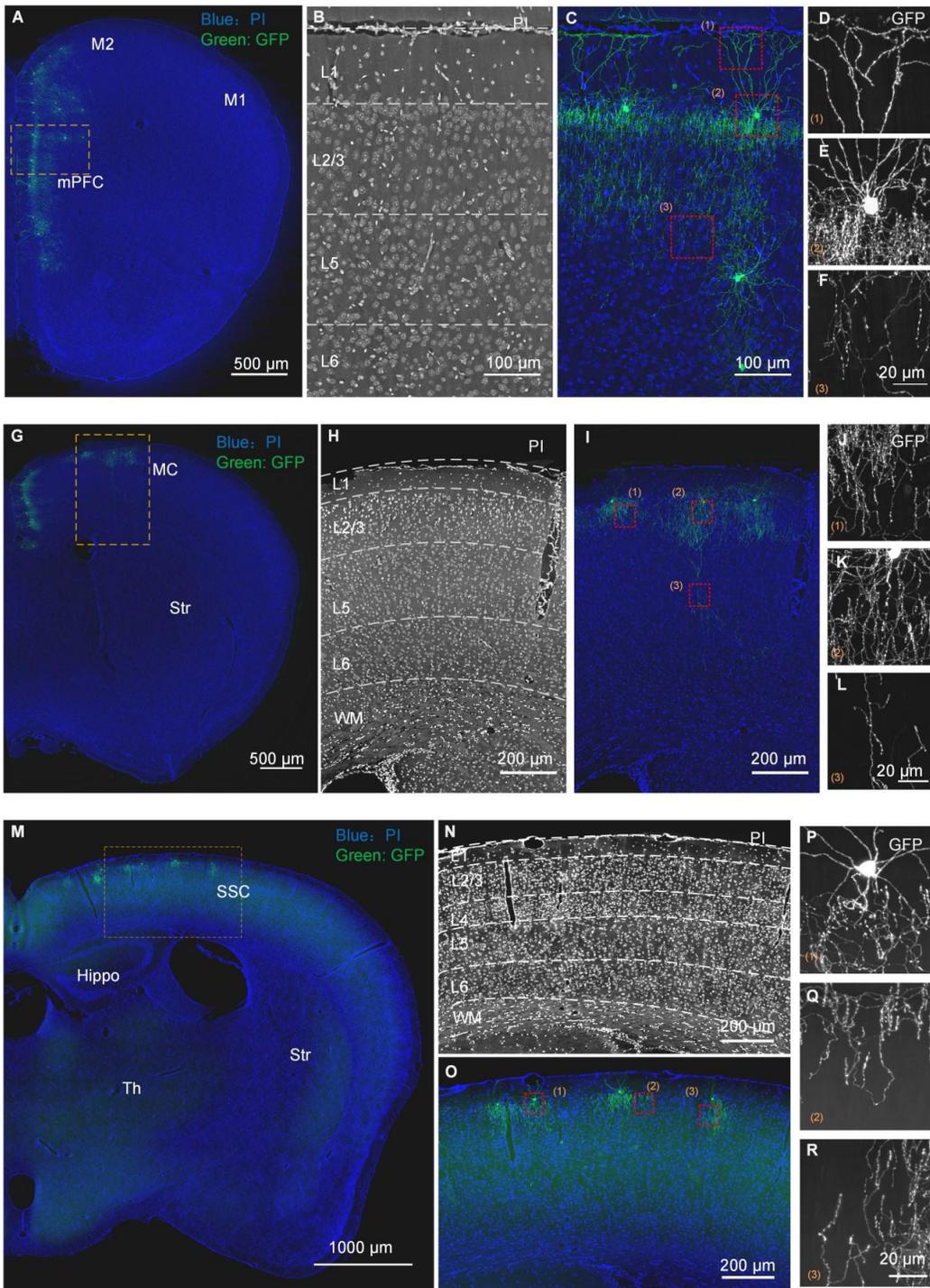


Figure A1S1. 1. Sparse labeling of AACs in mPFC, MC and SSC. Related to Figure A1.2. (A) Distribution of sparsely labeled AACs in mPFC shown as overlay of GFP channel (green, 200 μm max intensity projection) and PI (propidium iodide) channel (blue, 5 μm max intensity projection) images. (B-

C) PI channel (B) and overlay (C) images enlarged from the boxed region in (A). The cytoarchitecture was stained by PI dyes. Dashed lines indicate cortical layer boundaries. (D-F) High resolution GFP channel images show the fine structures of ACCs. Images are enlarged from boxed areas in (C). (G-L) Distribution of sparsely labeled AACs in MC. Legends are similar as those for (A-F). (M-R) Distribution of sparsely labeled AACs in SSC. Legends are similar as those for (A-F).

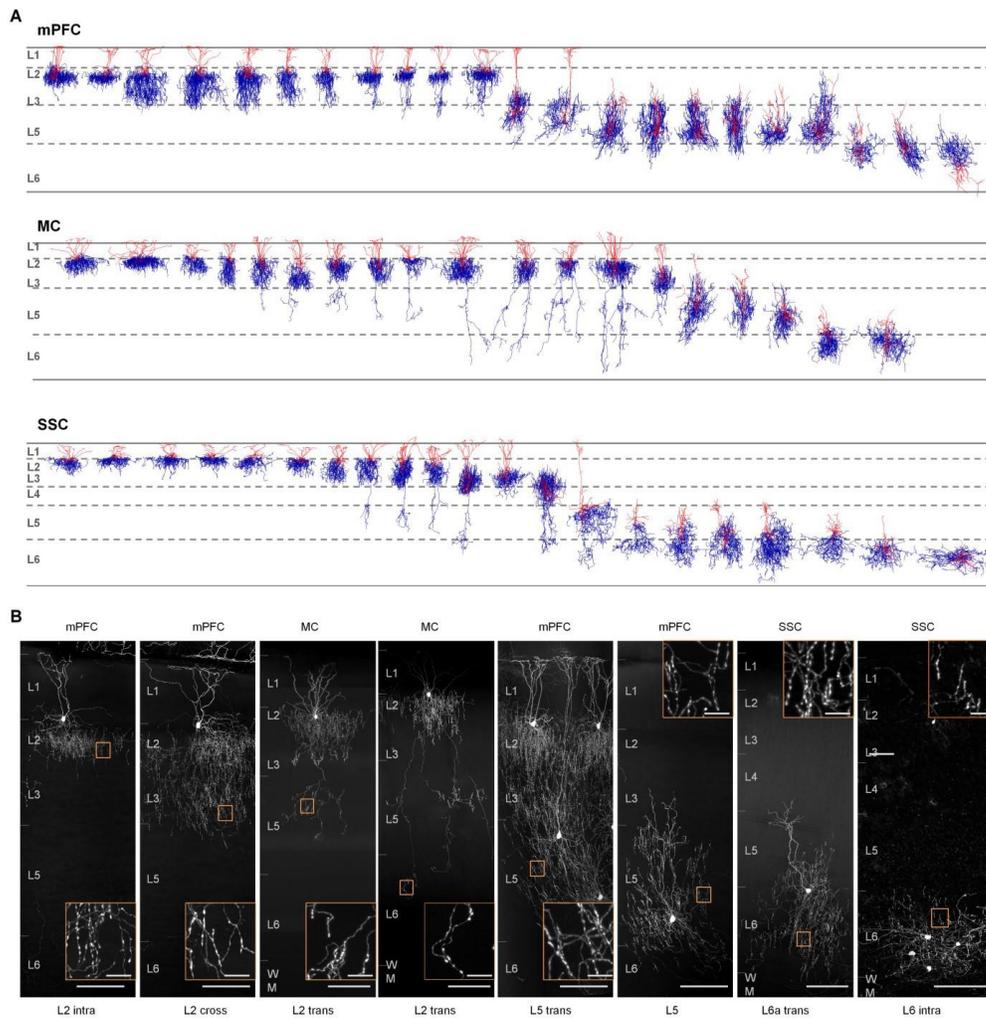


Figure A1S1. 2. AAC reconstructions (orientation adjusted) and typical AAC individuals. Related to Figure A1.3. (A) Full list of all reconstructed AACs in mPFC, MC and SSC (orientation readjusted). (B) Typical AAC subtypes in neocortex. 100 μm max intensity projection. Inserts are enlarged images from the boxes. Scale bars: 100 μm (large images) and 10 μm (inserts). (This figure corresponds to Figure A1.6 clustering analysis).

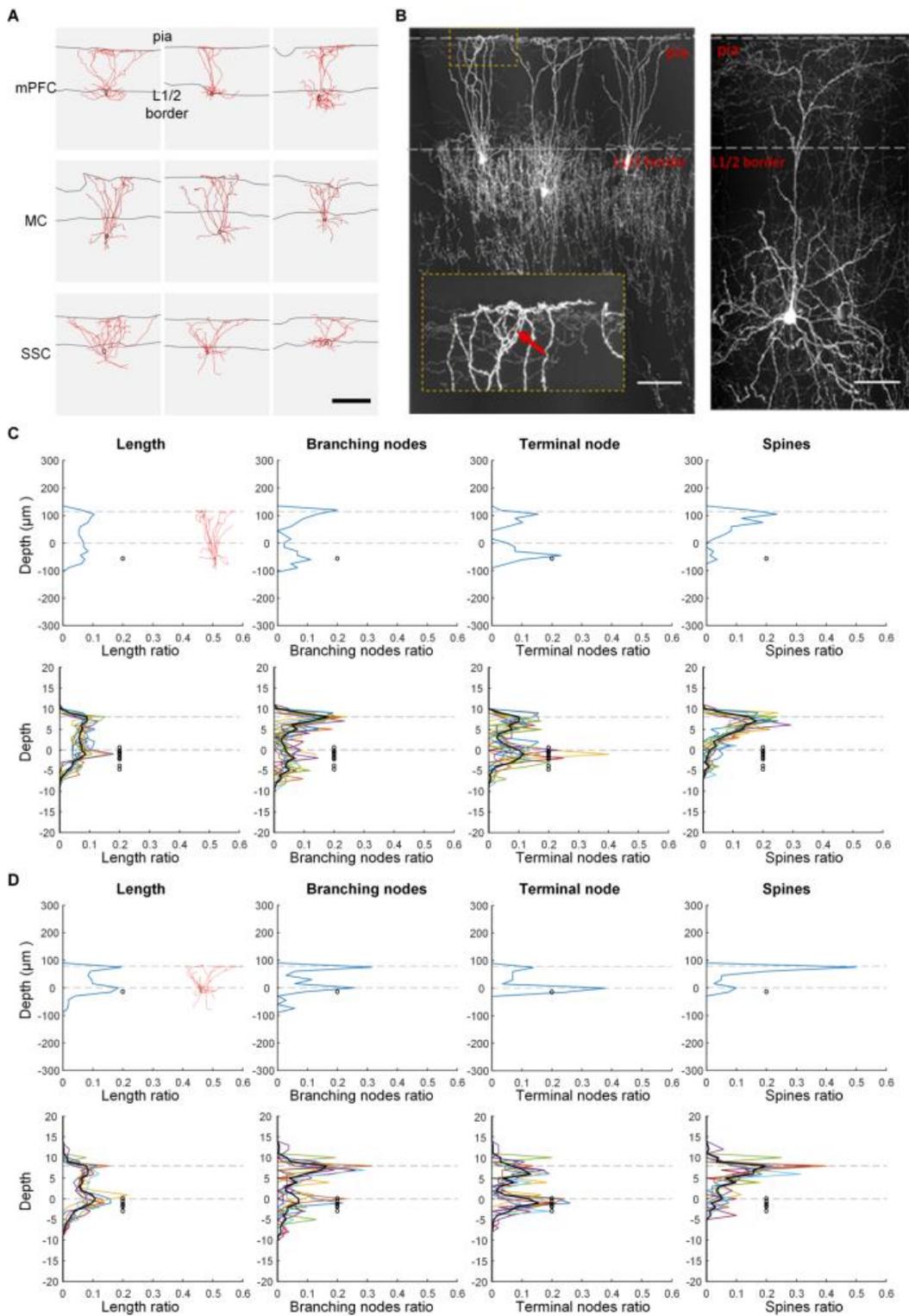


Figure A1S1. 3. Characteristics of L2 AAC dendrites. Related to Figure A1.4. (A) Representative reconstructions of AAC dendrites in mPFC, MC and SSC. Scale bar: 100 μm . Black circle: cell body. Red: dendrite. Gray lines indicate the positions of pia and L1/2 border. (B) Left: L2 AACs in mPFC. The apical dendrites of L2 AAC clamp the pia surface (box) and show lateral extensions (arrow). Right: the apical dendrites of a L3 pyramidal neuron do not contact pia and show different apical dendrite features compared with the dendrite of L2 AAC. Dashed lines indicate the pia and L1/2 border position. Scale bar: 100 μm . (C) Density plots of dendritic length, branching modes, terminal nodes and spines of L2 AACs in MC. Top row: density plots of a representative single cell. Bottom row: normalized density plots of the dendritic characteristics of all L2 AACs (n=10). Black curves: average of all the cells. Black circles: cell body positions in the coordinate. Dashed lines indicate the pia and L1/2 border. (D) Density plots of L2 AAC dendritic characteristics in SSC (n=7). Legends are similar as for (C).

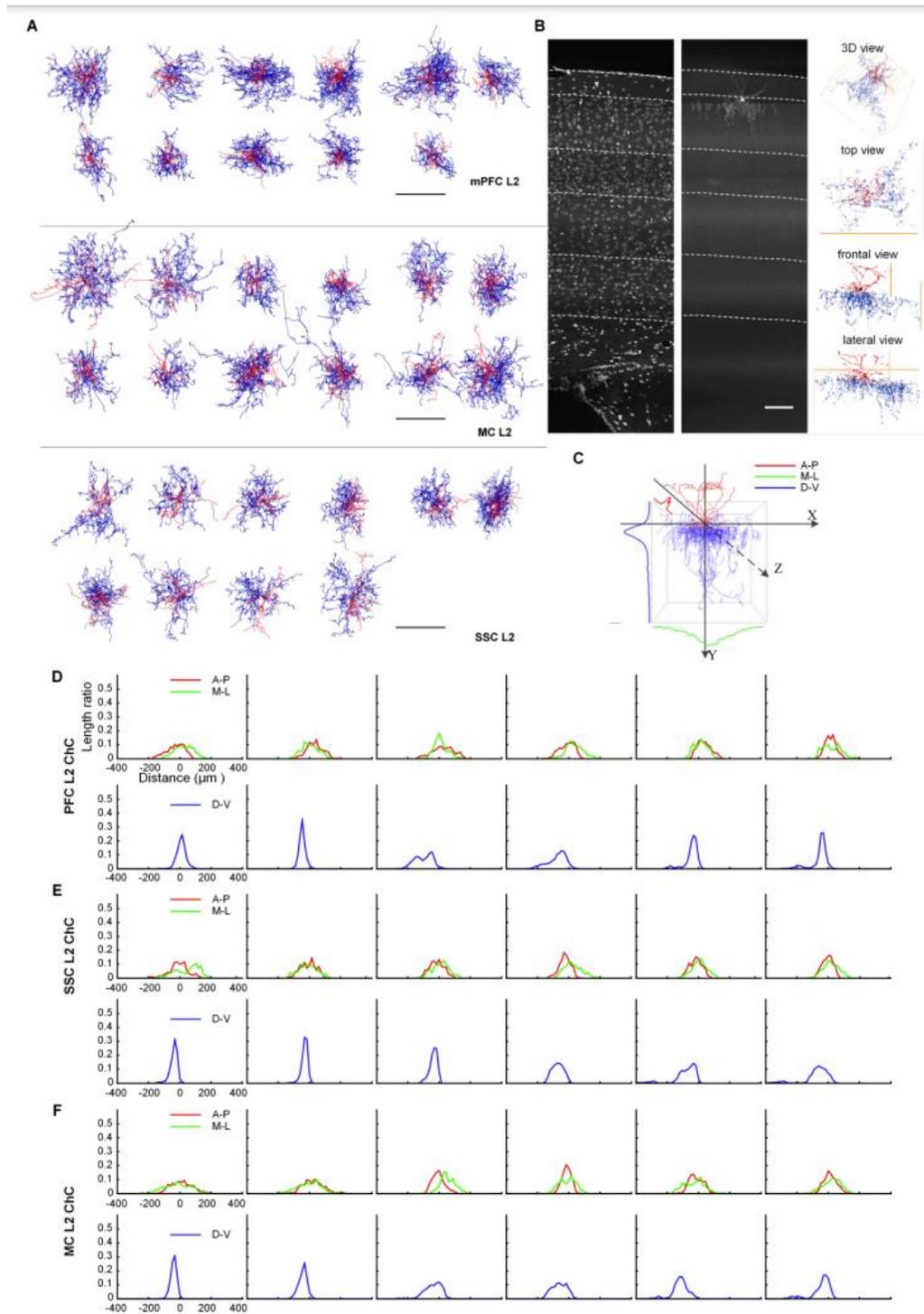


Figure A1S1.4. Cylindrical length density distribution pattern of L2 AAC axons. Related to Figure A1.5. (A) Top view projections of L2 AACs in mPFC, MC and SSC. Red: dendrites. Blue: axons. Black:

cell body. Scale bars: 200 μm . (B) 3D-view of a L2 AAC in SSC. Left: 5 μm max intensity projection of PI channel. Middle: 100 μm max intensity projection of GFP channel. Right: 3D view of cell reconstruction. Frontal view corresponds to coronal projection. Lateral view corresponds to sagittal projection. Scale bar: 100 μm . (C) A schematic of the projection of axon arbor distributions along anterior-posterior (A-P), medial-lateral (M-L) and dorsal-ventral (D-V) directions. (D-E) Axon length density distributions (ratio) of L2 AACs in mPFC (D), SSC (E) and MC (F). Top rows are length density curves along A-P and M-L axes. Bottom rows are length density curves along D-V axis. '0' point in x-axis corresponds to cell body position. Columns correspond to individual L2 AACs.

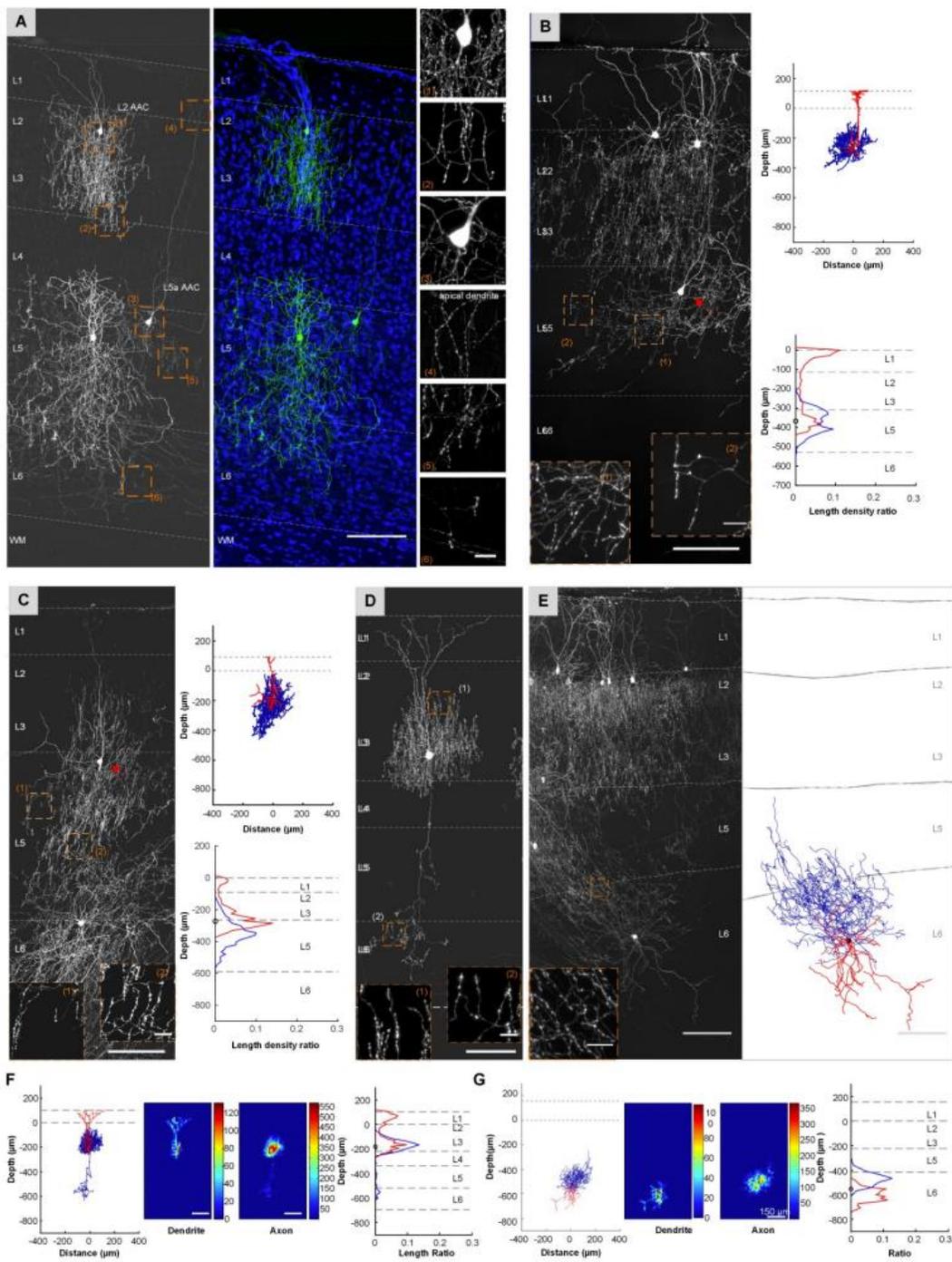


Figure A1S1.5. Representative individual and clustered AACs in the neocortex. Related to Figure A1.5. (A) Nearby L2 and L5a AACs in SSC. Left: 100 μm max intensity projection of GFP channel images. Middle: overlay of GFP channel and PI channel (5 μm max intensity projection) images. Right:

enlarged images from the boxes. Dashed lines: layer boundaries. Scale bars: 100 μm (left and middle) and 10 μm (enlarged images). Note the L5a AAC (cell #38) was incorrectly assigned to cluster 5 by the hierarchical clustering algorithm in Figure A1.6. (B-C) L5 AACs in mPFC (B) and MC (C). Left: 100 μm max intensity projection of GFP channel images. Scale bar: 100 μm . Right: Reconstruction (top) and length-density of dendrites (red) and axons (blue) along cortical depth (bottom). Inserts: enlarged from the boxes. Scale bar: 10 μm . Dashed lines: layer boundaries. Laminar borders are determined from the PI channel as in (A) (data not shown). Arrows indicate the cell bodies. (D) L3 AAC in SSC. 100 μm max intensity projection of GFP channel images. Inserts: enlarged images from the boxes. Dashed lines: layer boundaries. Scale bars: 100 μm and 10 μm (inserts). (E) Inverted L6 AAC in mPFC. Left: 200 μm max intensity projection of GFP channel images. Inserts: axon terminals (typical cartridges; enlarged from the box). Right: reconstruction. Gray lines indicate the layer boundaries. Scale bars: 100 μm and 10 μm (inserts). (F-G) length-density analysis of the AACs in (D) and (E). Left: reconstruction. Middle: heatmaps of length densities of dendrites and axons. Right: length-density of dendrites (red) and axons (blue) along cortical depth.

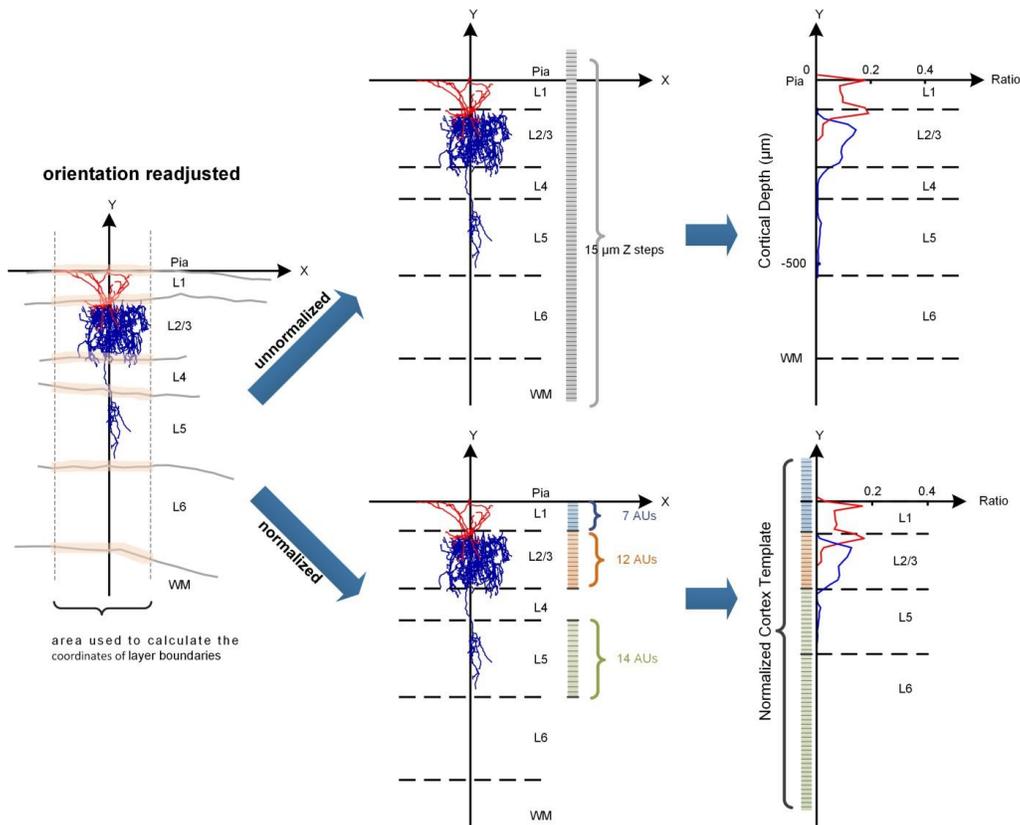


Figure A1S1. 6. Schematic of normalizing axonal and dendritic length-density distribution to a standard cortex template. Related to Figure A1.6. Note that L4 only exists in SSC, and L6-WM (white matter) boundary cannot be drawn in mPFC. Layer thicknesses vary significantly between areas.

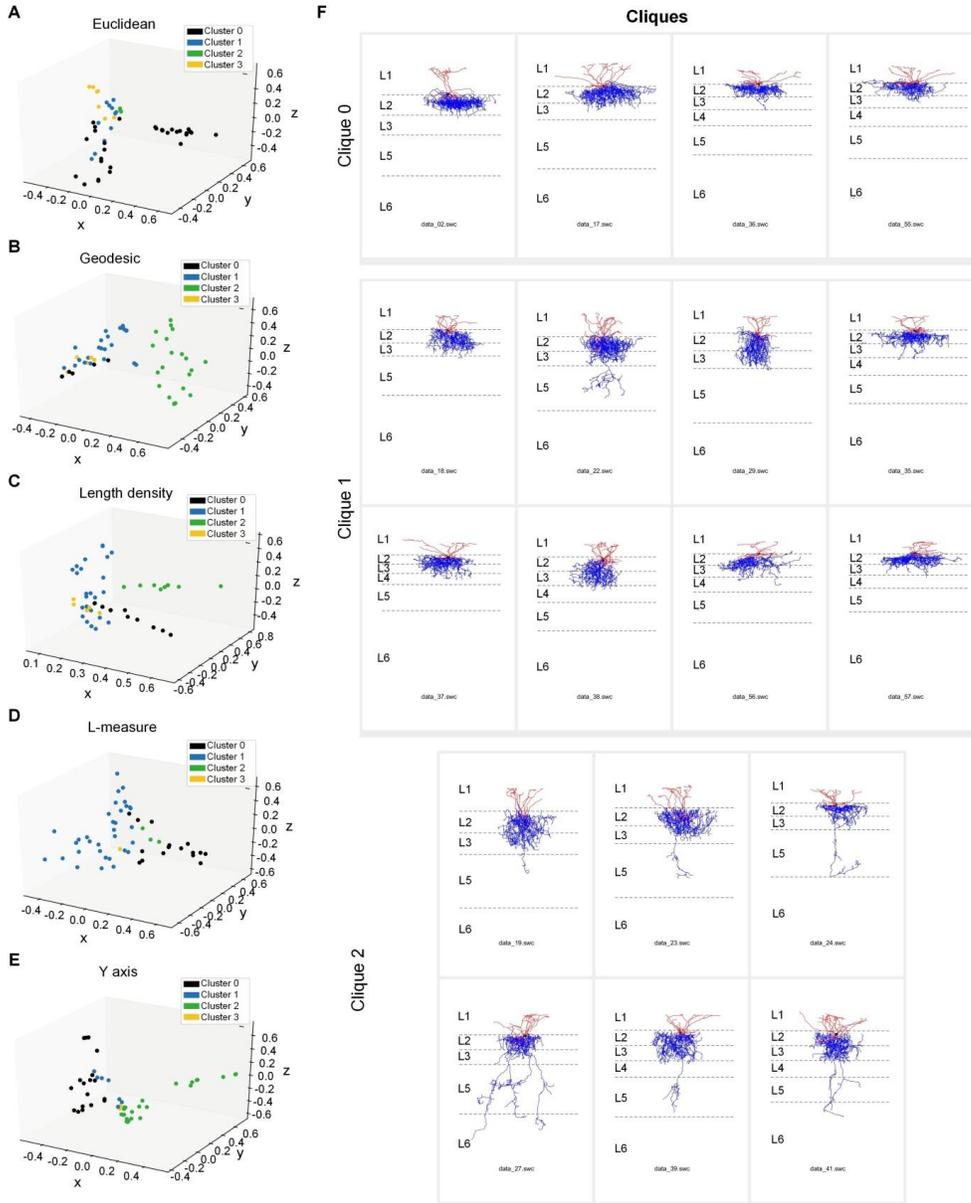


Figure A1S1. 7. Claque analysis of AAC types. Related to Figure A1.6 and STAR Methods. (A-E) The 3D Laplacian Eigenmap embeddings for each of the 5 metrics for $K=4$. (F) A full listing of the three cliques found from the robust clustering analysis. Dashed lines indicate the layer boundaries.

Table A1S1. 1. Complete cell list and Neurolucida dataset on morphometry.

Samples and Cells		Cell body					Dendrite										Axon											
Sample No.	Data No.	area	L1 thickness(µm)	layer	pa to L1/2border	soma_radius (µm)	Qty	Nodes	Ends	Full Length(µm)	Mean length (µm)	Max_branch_order	Branch Qty	Branch mean length(µm)	Branch length SD (µm)	Mean tortuosity	Tortuosity SD	Convex hull (µm ³)	Nodes	Ends	Full Length(µm)	Max_branch_order	Branch Qty	Branch mean length (µm)	Branch length SD	Mean tortuosity	Tortuosity SD	Convex hull (µm ³)
TDI21301010_cell3	1	PFC	143.9	L2	-69	5.2	7	38	46	3590.0	512.9	10	84	42.7	42.7	1.2	0.3	2749150	914	950	30068.6	30	1864	16.1	17.4	1.2	0.4	9095980
TDI21301017_cell2	2	PFC	126.3	L2	3	4.6	7	40	47	2281.0	325.9	9	87	26.2	30.6	1.2	0.2	2329810	552	555	17840.4	22	1107	16.1	18.3	1.2	0.4	5035430
TDI21301017_cell3	3	PFC	120.0	L2	5	5.3	7	36	43	2628.4	375.5	10	79	33.3	37.6	1.2	0.2	2180860	1059	1070	35480.7	28	2129	16.7	18.2	1.2	0.4	12164500
TDI21301013_cell2	4	PFC	121.2	L2	-20	6.4	6	48	57	3544.5	590.7	10	105	33.8	31.6	1.2	0.2	2628850	890	897	32588.6	31	1787	18.2	18.1	1.2	0.3	11142200
TDI21301013_cell3	5	PFC	130.8	L2	-2	5.8	6	35	41	3130.8	521.8	6	76	41.2	38.9	1.1	0.1	4314140	1033	1040	38993.4	31	2073	18.8	19.4	1.2	0.3	16858900
TDI21301012_cell3	6	PFC	136.2	L2	-26	5.4	10	19	29	2172.1	217.2	7	48	45.3	41.8	1.2	0.2	1548340	550	563	16449.9	30	1113	14.8	15.6	1.2	0.3	6089710
TDI21301012_cell6	7	PFC	117.9	L2	-14	4.0	5	20	28	1955.7	391.1	9	48	40.8	47.8	1.3	0.3	1633720	484	490	13552.8	39	974	13.9	17.5	1.3	0.6	5787130
TDI21301012_cell5	8	PFC	121.7	L2	-17	4.4	6	22	29	1782.3	297.1	8	51	34.9	29.8	1.2	0.2	1375690	372	383	14887.6	40	755	19.7	22.2	1.3	0.7	4881080
TDI21301008_cell4	9	PFC	126.8	L2	-2	4.6	6	31	37	2672.8	445.5	8	68	39.3	38.2	1.2	0.2	1896390	872	888	24561.9	50	1760	14.0	14.6	1.2	0.3	8062300
TDI21301008-cell13	58	PFC	124.8	L2	-23	5.3	5	63	71	2010.0	402.0	19	134	15.0	17.4	1.1	0.1	1372740	506	532	16630.4	39	1038	16.0	17.0	1.2	0.6	6840800
TDI21301012-cell4	59	PFC	142.7	L2	-2	4.4	6	25	31	1791.1	298.5	11	56	32.0	28.6	1.1	0.1	1806930	364	375	11865.6	25	739	16.1	15.2	1.3	0.6	3971610
TDI21301012_cell1	10	PFC	115.5	L5	-316	6.3	5	65	71	3554.1	710.8	14	136	26.1	34.1	1.2	0.2	5218500	936	956	31197.5	30	1892	16.5	17.7	1.2	0.5	21604200
TDI21301012_cell2	11	PFC	122.3	L5	-333	5.8	8	49	58	3652.5	456.6	9	107	34.1	40.8	1.1	0.1	3361780	1028	1040	34427.6	46	2068	16.6	16.7	1.3	0.6	13247400
TDI21301013_cell1	12	PFC	142.8	L5	-395	6.0	9	66	75	3713.1	412.6	12	141	26.3	27.2	1.1	0.1	3392320	1232	1248	34836.6	39	2480	14.0	15.3	1.2	0.4	18838400
TDI21301008_cell1	13	PFC	94.3	L5	-342	4.6	6	50	57	3701.2	616.9	9	107	34.6	39.8	1.2	0.1	3934560	879	895	29528.9	36	1774	16.6	17.9	1.2	0.3	12282400
TDI21302012_cell1	14	PFC	144.3	L5	-408	4.9	7	38	45	2858.0	408.3	10	83	34.4	36.6	1.1	0.1	3314500	618	622	19447.4	30	1240	15.7	15.9	1.2	0.4	9601130
TDI21301006_cell1	15	PFC	150.5	L5	-373	5.7	6	39	45	3661.1	610.2	10	84	43.6	51.4	1.2	0.1	5045700	897	898	35225.4	48	1795	19.6	21.3	1.3	0.4	19956300
TDI21301008-Cell2	52	PFC	141.9	L6	-454	4.8	7	33	41	3114.4	444.9	7	74	42.1	38.1	1.2	0.2	4152530	417	427	19901.2	20	844	23.6	27.6	1.2	0.3	14517700
TDI21301008-cell5	53	PFC	155.3	L6	-553	4.4	5	95	##	2826.1	565.2	18	199	14.2	17.6	1.2	0.2	2856580	477	478	13816.8	25	955	14.5	15.5	1.3	0.4	7347240
TDI21301008-Cell21	54	PFC	152.8	L6	-374	4.8	4	33	37	2989.2	747.3	10	70	42.7	37.8	1.2	0.1	3220960	573	582	21309.5	29	1155	18.4	20.5	1.3	0.3	9675610
TDI21301010_Cell	61	PFC	118.1	L5	-251	5.8	7	53	60	4164.6	594.9	12	113	36.9	37.8	1.1	0.1	5925090	638	640	21920.8	29	1278	17.2	19.2	1.3	0.5	13587400
TDI21301008_cell25	63	PFC	115.1	L5	-252	4.8	6	23	29	2409.0	401.5	10	52	46.3	53.4	1.1	0.1	2795260	484	486	15030.4	24	970	15.5	16.3	1.2	0.3	8248750
TDI21301017_cell5	16	MC	98.9	L2	7	4.4	5	42	47	4213.3	842.7	9	89	47.3	58.2	1.2	0.2	8192600	1232	1240	30864.7	34	2472	12.5	12.7	1.2	0.3	10372900
TDI21301017_cell10	17	MC	82.6	L2	2	3.8	6	30	36	2746.4	457.7	8	66	41.6	46.1	1.1	0.1	3151810	659	675	19797.1	26	1334	14.8	14.5	1.2	0.3	9437580
TDI21301012_cell12	18	MC	63.8	L2	-4	4.2	7	15	22	1875.9	268.0	5	37	50.7	51.1	1.2	0.2	1132300	558	572	13908.5	25	1130	12.3	11.6	1.2	0.3	3893400
TDI21301008_cell14	19	MC	114.5	L2	-55	5.0	4	39	45	2791.2	697.8	9	84	33.2	38.9	1.2	0.3	1755900	500	513	15678.6	48	1013	15.5	14.6	1.2	0.3	7608080
TDI21301015_cell3	20	MC	95.1	L3	-124	4.1	5	29	34	2618.8	523.8	8	63	41.6	43.5	1.2	0.2	1899200	566	569	17009.2	25	1135	15.0	15.8	1.3	0.4	6355410
TDI21302012_cell2	21	MC	111.9	L2	-13	4.5	6	30	36	2807.8	468.0	7	66	42.5	39.8	1.2	0.2	2566650	510	511	16752.9	25	1021	16.4	16.2	1.2	0.3	9892980
TDI21302012_cell4	22	MC	101.4	L2	-29	3.5	5	36	41	3218.6	643.7	7	77	41.8	43.5	1.2	0.2	3172040	734	736	20255.8	28	1470	13.8	14.8	1.2	0.4	9239210
TDI21301020_Cell2	23	MC	96.7	L2	-56	4.5	8	54	62	3016.5	377.1	12	116	26.0	44.1	1.2	0.3	3560520	646	647	17178.9	26	1293	13.3	13.9	1.2	0.3	10705900
TDI21301012_cell8	24	MC	71.7	L2	-7	3.4	8	26	34	1942.0	242.7	7	60	32.4	31.3	1.2	0.1	1392880	467	472	10413.7	27	939	11.1	13.6	1.2	0.3	8574650
TDI21301017_cell8	25	MC	97.3	L2	-13	5.0	6	44	50	4361.7	727.0	8	94	46.4	43.8	1.2	0.2	4282790	969	980	28597.2	42	1949	14.7	15.0	1.2	0.4	31085400
TDI21301019_cell3	26	MC	97.4	L2	-23	4.4	7	54	61	3491.2	498.7	18	115	30.4	36.8	1.2	0.2	3076410	647	654	19125.0	43	1301	14.7	17.7	1.2	0.3	30578100

TDI21301020_Cell1	27	MC	103.4	L2	-3	5.2	8	29	37	2981.0	372.6	7	66	45.2	45.7	1.2	0.2	3216450	532	537	15175.3	25	1069	14.2	20.6	1.2	0.3	19408000
TDI21301009_Cell1	28	MC	117.4	L2	-53	4.6	6	30	36	4653.8	775.6	6	66	70.5	69.4	1.2	0.1	7167070	1133	1137	36603.5	43	2270	16.1	17.4	1.3	0.4	39961000
TDI21301018_cell3	29	MC	80.9	L2	-18	4.0	8	17	25	1849.0	231.1	6	42	44.0	36.2	1.3	0.3	1184710	528	532	13860.8	29	1060	13.1	13.3	1.2	0.3	4284460
TDI21301008_Cell20	30	MC	100.5	L6	-499	5.6	5	46	51	3059.7	611.9	10	97	31.5	31.5	1.2	0.1	3672310	873	882	24464.7	28	1755	13.9	15.5	1.2	0.3	15156700
TDI21301012_Cell11	31	MC	74.4	L6a	-316	4.7	4	57	61	2706.4	676.6	16	118	22.9	27.7	1.2	0.3	2222030	722	729	27718.2	34	1451	19.1	21.8	1.3	0.4	13643500
TDI21301015_Cell2	32	MC	115.6	L5	-358	5.2	7	38	45	2978.4	425.5	6	83	35.9	33.2	1.2	0.1	2596490	668	668	22686.6	20	1336	17.0	19.6	1.2	0.3	16156600
TDI21301019_Cell1	33	MC	77.2	L5	-297	4.8	4	81	88	2654.1	663.5	23	169	15.7	23.3	1.1	0.2	2027960	459	462	16046.3	19	921	17.4	18.3	1.2	0.3	7262440
TDI21301017_Cell1	34	MC	91.3	L5a	-178	4.9	7	31	38	2588.2	369.7	9	69	37.5	39.9	1.1	0.1	4791730	949	954	27735.1	30	1903	14.6	16.3	1.2	0.4	14860400
TDI21301013_Cell5	35	SSC	66.0	L2	-5	4.5	9	23	32	1899.2	211.0	7	55	34.5	29.1	1.1	0.1	1077580	628	644	13993.5	33	1272	11.0	11.7	1.2	0.3	8013190
TDI21301017_Cell12	36	SSC	71.3	L2	1	3.6	9	49	58	2051.9	228.0	16	107	19.2	25.7	1.1	0.2	1615340	575	581	12397.4	26	1156	10.7	10.9	1.2	0.3	3965010
TDI21301012_Cell19	37	SSC	68.4	L2	-10	4.9	6	25	31	2019.5	336.6	10	56	36.1	40.5	1.2	0.2	2357480	548	558	13722.2	32	1106	12.4	12.7	1.2	0.4	4112270
TDI21301017_Cell13	38	SSC	72.5	L2	-16	4.2	3	47	50	2575.3	858.4	10	97	26.6	31.1	1.1	0.2	1939460	456	458	13377.0	22	914	14.6	15.7	1.2	0.3	4594670
TDI21301017_Cell6	39	SSC	78.9	L2	-13	3.8	6	29	35	2134.1	355.7	8	64	33.4	32.9	1.1	0.2	1887100	529	537	15800.1	24	1066	14.8	15.3	1.2	0.2	8162120
TDI21301018_Cell1	40	SSC	84.7	L2	-31	4.7	7	46	53	3130.4	447.2	10	99	31.6	37.4	1.2	0.3	2037070	889	893	23004.7	30	1782	12.9	12.2	1.2	0.2	9603060
TDI21301008_Cell19	41	SSC	82.2	L2	-19	5.7	6	31	38	2989.7	498.3	7	69	43.3	47.6	1.2	0.2	2340220	475	481	15363.1	21	956	16.1	16.2	1.2	0.3	11465400
TDI21301010-cell7	55	SSC	64.6	L2	-5	4.2	6	34	40	2154.9	359.1	8	74	29.1	30.8	1.2	0.2	1482180	382	390	12617.0	30	772	16.3	17.2	1.3	0.5	4625740
TDI21301012-cell10	56	SSC	55.1	L2	0	4.4	7	21	28	1873.5	267.6	8	49	38.2	30.9	1.1	0.2	1904680	532	546	11988.4	45	1078	11.1	11.0	1.2	0.8	5513540
TDI21301012-cell13	57	SSC	56.0	L2	3	4.6	6	31	37	2120.8	353.5	8	68	31.2	37.0	1.2	0.1	1807520	588	597	14474.3	48	1185	12.2	12.3	1.2	0.3	5334430
TDI21301009_Cell2	42	SSC	100.9	L3	-179	4.9	6	33	39	3809.1	634.8	6	72	52.9	46.9	1.2	0.2	7081040	1150	1157	29169.6	34	2307	12.6	13.6	1.2	0.3	15866800
TDI21302012_Cell3	43	SSC	113.8	L3	-108	4.2	5	45	51	3087.0	617.4	10	96	32.2	37.7	1.2	0.2	2548630	476	480	16718.3	31	956	17.5	16.9	1.2	0.3	8351940
TDI21301017_cell11	44	SSC	78.7	L4	-177	5.5	8	21	29	2759.8	345.0	6	50	55.2	53.8	1.1	0.1	3505300	1049	1052	31759.6	25	2101	15.1	16.9	1.2	0.3	18153000
TDI21301014_cell2	45	SSC	104.2	L5	-320	4.8	6	45	51	3747.8	624.6	11	96	39.0	43.2	1.2	0.2	8996910	669	672	23437.1	35	1341	17.5	18.9	1.2	0.3	19764300
TDI21301013_cell6	46	SSC	53.4	L5	-253	4.5	3	19	22	1561.1	520.4	7	41	38.1	34.1	1.1	0.1	1760840	325	332	13054.5	25	657	19.9	23.9	1.2	0.4	10309900
TDI21301017_cell15	47	SSC	75.0	L5b	-426	5.3	8	52	60	2751.3	343.9	12	112	24.6	26.3	1.1	0.1	2885760	1239	1242	39214.2	40	2481	15.8	17.9	1.2	0.3	27046400
TDI21301018_cell2	48	SSC	84.9	L6a	-320	4.8	6	28	34	2747.9	458.0	7	62	44.3	39.9	1.2	0.1	2369140	712	717	20208.3	33	1429	14.1	14.8	1.2	0.3	8647810
TDI21301015_cell1	49	SSC	67.9	L6a	-463	5.5	9	62	71	3339.1	371.0	17	133	25.1	24.0	1.2	0.1	4231820	709	712	25020.3	25	1421	17.6	19.3	1.3	0.4	19235300
TDI21301010_cell6	50	SSC	64.8	L6a	-306	4.1	7	23	31	2136.2	305.2	5	54	39.6	34.3	1.2	0.2	1327990	253	259	16517.3	47	512	32.3	37.9	1.4	0.4	11768100
TDI21301010_cell5	51	SSC	50.5	L6	-334	4.4	9	24	33	2582.8	287.0	5	57	45.3	41.8	1.1	0.1	2537150	485	495	22684.6	23	980	23.1	26.6	1.2	0.4	14605800
TDI21301009_cell5	62	SSC	78.1	L6	-549	5.4	8	18	26	2226.6	278.3	5	44	50.6	41.9	1.1	0.1	2675880	389	385	16105.4	24	774	20.8	23.0	1.2	0.4	8120620

APPENDIX 2: AN OPEN-SOURCE FRAMEWORK FOR NEUROSCIENCE METADATA MANAGEMENT APPLIED TO DIGITAL RECONSTRUCTIONS OF NEURONAL MORPHOLOGY

Kayvan Bijari¹, Masood A. Akram¹ and Giorgio A. Ascoli^{1*}

¹Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA (USA)

*This appendix was published in Brain Informatics. 7, 2 (2020).

Abstract

Research advancements in neuroscience entail the production of a substantial amount of data requiring interpretation, analysis, and integration. The complexity and diversity of neuroscience data necessitate the development of specialized databases and associated standards and protocols. NeuroMorpho.Org is an online repository of over one hundred thousand digitally reconstructed neurons and glia shared by hundreds of laboratories worldwide. Every entry of this public resource is associated with essential metadata describing animal species, anatomical region, cell type, experimental condition, and additional information relevant to contextualize the morphological content. Until recently, the lack of a user-friendly, structured metadata annotation system relying on standardized terminologies constituted a major hindrance in this effort, limiting the data release pace. Over the past 2 years, we have transitioned the original spreadsheet-based metadata annotation system of NeuroMorpho.Org to a custom-developed, robust, web-based framework for extracting, structuring, and managing neuroscience information. Here we release the metadata portal publicly and explain its functionality to enable usage by data contributors. This framework facilitates metadata annotation, improves

terminology management, and accelerates data sharing. Moreover, its open-source development provides the opportunity of adapting and extending the code base to other related research projects with similar requirements. This metadata portal is a beneficial web companion to NeuroMorpho.Org which saves time, reduces errors, and aims to minimize the barrier for direct knowledge sharing by domain experts. The underlying framework can be progressively augmented with the integration of increasingly autonomous machine intelligence components.

Keywords: Neuroscience curation, Metadata extraction, Knowledge engineering, Data sharing, Information management tools, Neuronal morphology

Introduction

Neuroscience is continuously producing an immense amount of complex and highly heterogeneous data typically associated with peer-reviewed publications. When building data-driven models of brain function, computational neuroscientists must engage in the laborious task of reviewing, annotating, and deriving many parameters required for numerical simulations. More generally, the process of curation consists of extracting, maintaining, and adding value to digital information from the literature and underlying datasets (Bandrowski et al., 2012). Mature reference management tools exist to aid general-purpose bibliography organization and content annotation, including Zotero (Puckett, 2011), Mendeley (Holt Zaugg, Richard E. West, Isaku Tateishi, Daniel L.

Randall, 2011), and EndNote (Agrawal & Rasouli, 2019). Moreover, community-sourced terminologies (Bug et al., 2008; Gardner et al., 2008; Hamilton et al., 2017; Shepherd et al., 2019) and domain-specific markup languages (Gleeson et al., 2010; Goddard et al., 2001; Grewe et al., 2011) provide human-interpretable controlled vocabularies and machine-readable file formats, respectively. Efforts are also underway to generate standardized data models (Gleeson et al., 2019; Rübél et al., 2015; Teeters et al., 2015) and to formalize related concepts into robust ontologies (Hamilton et al., 2012; Koopmans et al., 2019; Larson & Martone, 2013). As a result, full-text information retrieval systems are becoming indispensable research aids (Falagas et al., 2008; Hutchins et al., 2019; Müller et al., 2008, 2018).

Despite promising progress, neuroscience and related fields lacked until recently a user-friendly tool to annotate a dataset or journal article across a customizable variety of fields with a set of controlled vocabularies. At the same time, a systematic and well-documented extraction process is essential to keep the curated metadata updated over time and portable between different projects (O'Reilly et al., 2017). Perhaps the sole example of an open-source, web-based framework for the acquisition, storage, search, and reuse of scientific metadata is the CEDAR workbench (Gonçalves et al., 2017). On the one hand, the entirety of neuroscience is too broad and diverse to fully benefit from an all-encompassing metadata annotation tool. On the other, the most useful motivating applications are typically task specific and, consequently, difficult to compare with other developed tools. Meanwhile, several fundamental metadata dimensions, including details

about the animal subject, the location within the nervous system, and the experimental condition, are largely common to even considerably distinct subfields of neuroscience. One possible approach is therefore to design a practical solution to a specific problem of interest while adhering to a strictly open-source implementation that may foster broad adoption and custom adaptation throughout the neuroscience community.

Here, we introduce a resource developed to promote and facilitate data sharing and metadata annotation for NeuroMorpho.Org, a repository providing unrestricted access to digital reconstructions of neuronal and glial morphology (Akram et al., 2018b; Ascoli et al., 2007a). The acquisition and release of morphological tracings begin with the continuous identification of newly published scientific reports describing data of interest (Halavi et al., 2012a, 2012a; Maraver et al., 2019b). To annotate the reconstructions with proper metadata, the repository administrators have also been inviting data contributors to provide suitable information through a semi-structured Excel spreadsheet (Parekh et al., 2015a). While the ecosystem of neuronal reconstructions has coalesced around a simple data standard for over two decades (Nanda et al., 2018b), selection and interpretation of metadata concepts remain highly variable and inconsistent. Thus, for every new dataset, a team of trained curators must validate or reconcile the author-provided information, complemented as needed by the associated publication, with the metadata schema and preferred nomenclature of the database. Many data releases also introduce new metadata concepts, which need to be integrated into the existing ontology and require updating relevant database hierarchies with appropriate

terms. Although the described process is time-consuming, labor-intensive, and error-prone, metadata annotation is instrumental to enable NeuroMorpho.Org semantic queries (Polavaram & Ascoli, 2017a) and machine accessibility through Application Programming Interfaces (Ascoli et al., 2017c).

This article presents the NeuroMorpho.Org metadata portal, a novel, open-source, web-based tool for the efficient annotation and collaborative management of data descriptors for digital reconstructions of neuronal and glial morphology. The main goal of this effort is the gradual automation of the metadata extraction process to reduce the burden on database curators, thus streamlining the data release workflow for the benefit of the entire research community. A related motivation is to bring domain expertise closer to the crucial task of metadata curation by empowering data contributors with direct dataset annotation through a graphical user interface. The longer-term vision is to lay the training data foundation for augmenting neuro-curation with semi-autonomous machine learning components such as recommendation systems or natural language processing tools (Benedetti et al., 2019; Bijari, Zare, et al., 2020; Egyedi et al., 2018). With this report, we freely release the documented code base to date and welcome modifications or improvements by other developers to tailor the metadata management platform for different neuroscience initiatives.

Methods

The metadata portal is designed to match the NeuroMorpho.Org metadata structure. Here first we summarize the organization of reconstruction metadata in this resource and then explain how the architectural design of the portal optimally serves the needs of the project.

Organization of NeuroMorpho.Org metadata

NeuroMorpho.Org stores over 120,000 digital reconstructions of neuronal and glial morphology from nearly 650 independent laboratories and more than 1000 peer-reviewed articles. Each reconstruction is associated with detailed metadata across 25 dimensions thematically grouped into five different categories, namely *animal*, *anatomy*, *completeness*, *experiment*, and *source* (Parekh et al., 2015a).

The animal category specifies the subject of the study: species, strain, sex, weight, development stage, and age.

The anatomy category designates the brain region and cell type. Each of these two dimensions is hierarchically divided into three levels, from generic to specific: for instance, hippocampus/CA1/pyramidal layer and interneuron/basket cell/parvalbumin-expressing. Three considerations are especially important in this regard: first, additional information can be added in multiple entries at the third level. In the above example, the

brain region could be further annotated as left and dorsal; and the cell type as fast-spiking and radially oriented. Second, the anatomical hierarchies are loosely rather than strictly organized since the specific details reported in (and relevant for) different studies vary considerably. If another paper describes the brain region of its dataset simply as dorsal hippocampus (without mentioning sub-area and layer), the concept “dorsal” would shift up to the second level. Third, both brain regions and cell types depend dramatically on the animal species, and most substantially diverge at the vertebrate vs. invertebrate taxa. Whenever possible, NeuroMorpho.Org follows the BrainInfo classification and NeuroNames terminology for vertebrates (Bowden et al., 2012), and Virtual Fly Brain for invertebrates (Osumi-Sutherland et al., 2012).

The completeness category provides details on the relative physical integrity of the reconstruction (accounting for tissue sectioning, partial staining, limited field of view, etc.), the structural domains included in the tracing (soma, axons, dendrites, undifferentiated neurites or glial processes), and the morphological attributes included or excluded from the measurement (most importantly, diameter and the depth coordinate).

The experiment category consists of methodological information describing the preparation protocol (e.g. in vivo, slice or culture), condition (control vs. lesioned, treated or transgenic), visualization label or stain, thickness and orientation of slicing or optical sections, objective type and magnification, tissue shrinkage and eventual corrections, and the tracing software.

The fifth category, source, provides details on the contributing laboratory, the reference publication, the original digital file formats, and the dates of receipt and release.

If any metadata dimension is not returned by the author or mentioned in the publication, the corresponding entry is marked as “Not reported” in the repository.

Here we refer to ‘dataset’ as a collection of reconstructions associated with a single peer-reviewed publication. Many datasets are naturally divided into distinct metadata groups, either as a focus of the study (e.g. control vs. experimental condition) or because of cell-level specification of a particular variable (often animal sex or age). Typically, almost all metadata features are identical across the entire dataset except for specific details varying between groups. NeuroMorpho.Org preserves the same annotation organization at the levels of dataset, groups, and individual cells (Fig. A2.1). This intuitive yet compact structure conveniently allows both comparative statistical analyses and machine-readable accessibility via APIs.

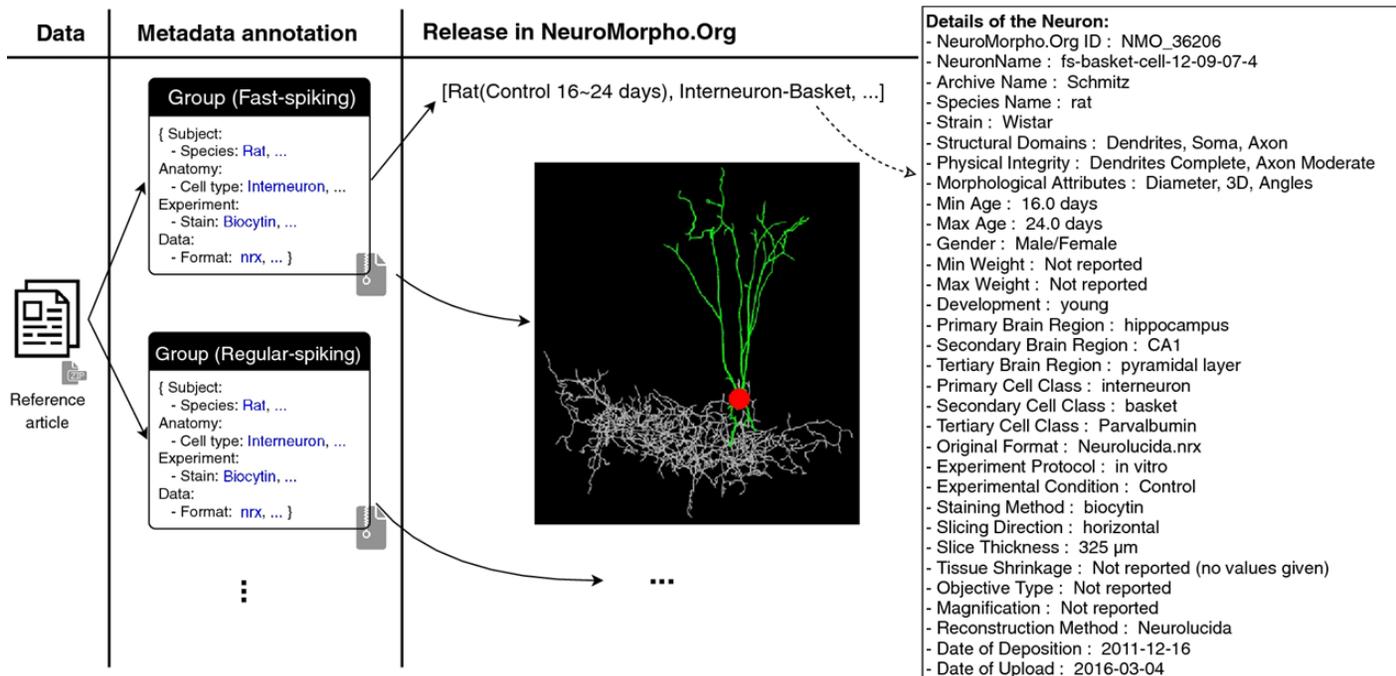


Figure A2. 1. Metadata organization in NeuroMorpho.Org. Every dataset is associated with a publication and is typically divided into homogeneous annotation groups, each containing several reconstructed cells. Dots in the figure indicates continuation of groups and reconstructions. The formal database schema is publicly available at neuromorpho.org/images/Schema.png

Design and implementation of the metadata portal

To ensure flexibility, scalability, portability, and efficiency, the metadata portal is designed based on the model-view-controller (MVC) software architecture (Bass et al., 2003). This modular approach separates the application into three essentially independent components. The *model* represents the metadata structure and reflects the constraints, relations, and formats stored in the database through an object-relational mapper (ORM). The *view* defines the display presented to the operator through the graphical user interface (GUI). The *controller* mediates the requests of the user, interacts with the model, and generates an appropriate response for the view (Fig. A2.2). While anchoring the architectural foundation of the metadata portal onto a safe and trusted design pattern, the novelty of this development mostly lies in its goal and features that assist users in the metadata curation process.

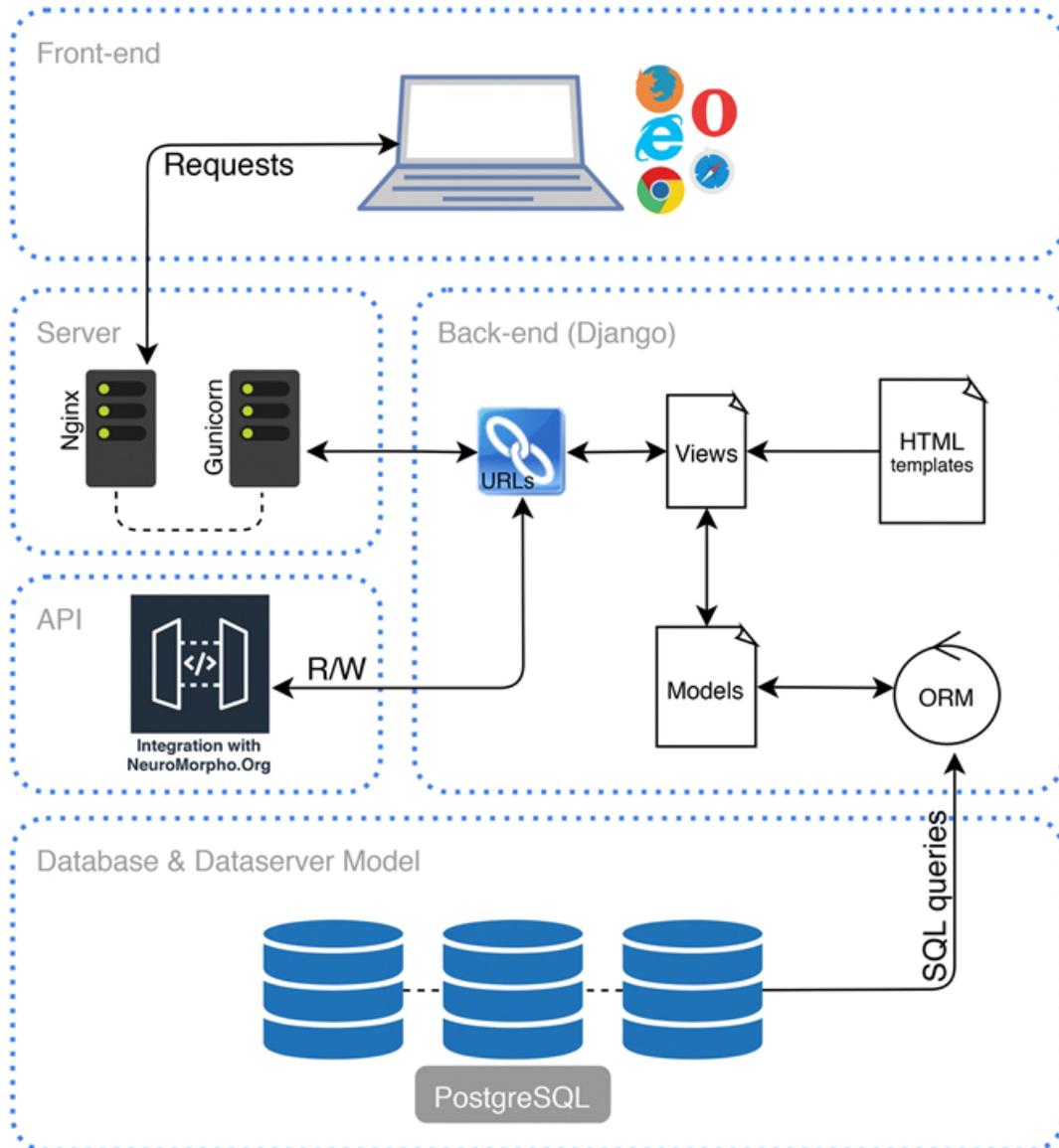


Figure A2. 2. Overview of the system’s architecture. The code base of the metadata portal is running on Nginx and Gunicorn webservers. The Django controller handles all requests submitted by the users or received through the application programming interface (API), translates them into machine-readable commands and database queries, and returns the proper results

The entire implementation abides by open-source principles and relies solely on open-source resources. The relational models of the portal in addition to the data are maintained in PostgreSQL, a fast, secure, and extensible relational database management system. The user interface is formulated by HTML, JavaScript, and Bootstrap, a Cascading Style Sheet (CSS) framework directed at responsive front-end web development. The control back-end is programmed in Django, a Python-based framework emphasizing pluggable and reusable elements, to regulate the interactions between database and users. Such modular yet integrated web-based framework offers rapid, cost-effective, and customizable application development. The resulting application is effortlessly accessible anytime across different platforms, enhancing interoperability and enabling different classes of users (authors, admins, and curators) to use the system independently while maintaining their work in the database.

The metadata portal encompasses most of the essential components to fulfill the curation needs of NeuroMorpho.Org. At the same time, it is also continuously evolving as new operational capabilities are prioritized. Recently developed features include: (i) the API (<http://cng-nmo-meta.orc.gmu.edu/api/>) enabling data interaction between the metadata portal and NeuroMorpho.Org; (ii) keyword search (<http://cng-nmo-meta.orc.gmu.edu/search/>), a user-friendly search engine allowing users to look for available terms in the database and their hierarchy; and (iii) bulk-modification feature, providing the ability to modify a large portion of terms within datasets.

The user interface of metadata portal offers seamless access to different parts and features of the system. The main page (<http://cng-nmo-meta.orc.gmu.edu/>) lists all active datasets. Each dataset is annotated with the name of the data contributor, publication identifiers (PMID and URL), and information regarding grant support. Metadata groups and their corresponding labels can be entered manually or are automatically created upon uploading grouped reconstruction files. Next, users select the actual entries for every metadata dimension, and the entire information remains accessible and editable through the web form. A detailed step-by-step metadata annotation protocol follows at the end of the Results.

Results

We deployed the metadata portal for internal usage in the NeuroMorpho.Org curation team in spring 2018 after release v.7.4 of the database, which contained 86,893 reconstructions. The most recent release at the time of this writing (fall 2019), v.7.9, contains 121,578 reconstructions. Thus, we completed five full releases and annotated nearly 35,000 new reconstructions using the novel system described in this article. Moreover, we analyzed the records regarding metadata entry over four releases prior to deployment of the current system, namely, from right after release v7.0 (fall 2016), which contained 50,356 reconstructions. In the next section, we describe the positive impact on the project of switching from offline spreadsheet annotation to the web-based metadata portal.

Metadata complexity, time saving, and error reduction

The metadata form in NeuroMorpho.Org employs more than 40 fields to encompass the details of the experiment, as several dimensions (e.g. animal weight and age) require more than one field (e.g. a numerical value and a unit scale). If treated as free text entry, many terms can be written in multiple equivalent variants, as in ‘mouse’, ‘Mouse’, ‘mice’, ‘mus musculus’ as well as being prone to semantically deviant typos (‘moose’). When considering the combination of all metadata fields, even in the absence of errors, the exact same information can be annotated in more than 10,000 distinct ways. Such an extreme case of combinatorial synonymy raises serious database management issues, in addition to slowing down search queries and requiring substantially inflated curation efforts. While the ‘mouse’ example may appear innocuous, even professional annotators can rapidly slide outside their zone of comfort when trying to distinguish between terminological equivalence and subtle but important differences in a genetic manipulation, staining process or electrophysiological firing pattern. The metadata portal offers a solution based on a corpus of controlled vocabularies consisting of public NeuroMorpho.Org content practically organized in user-friendly dropdown menus with autocomplete functionality and ‘similar hits’ suggestions. Moreover, the web form is endowed with hierarchical logic so that, for example, rat strains are not presented if mouse is selected as species.

Another major aspect of metadata annotation is the ongoing necessity to add new terms to describe previously unencountered entries. While certain dimensions, such as developmental stage, sex, objective type, and physical integrity, remain essentially

unaltered over time, others, including brain regions, cell types, and experimental conditions, grow continuously at rates of approximately 5% (amounting to hundreds of new entries) per database release (Table A2.1). The web-based system facilitates the management of new concepts by enabling submission of free-text entries when needed; these are logged in real time into the database, allowing secondary review and provenance tracking.

Table A2. 1. Number of distinct scientific concepts in the metadata portal arranged by category, with separate indication of newly added concepts in recent releases.

Category	Dimension	Concepts in v.7.0	New in 7.1	New in 7.2	New in 7.3	New in 7.4	New in 7.5	New in 7.6	New in 7.7	New in 7.8	New in 7.9	Concepts in v.7.9	Increase per release
Animal	Species	32	5	1	1	13	4	1	3	3	2	65	5.23 %
	Strain	218	31	35	0	102	26	12	37	37	28	526	6.43 %
	Others ^a	10	1	0	0	0	0	0	0	0	0	11	0.91 %
Anatomy	Brain region	1473	48	34	41	82	15	11	114	48	195	2061	2.94 %
	Cell type	972	62	59	45	99	44	10	93	99	645	2128	5.75 %
Experiment	Staining methods	3	4	8	7	6	2	3	7	14	6	60	9.83 %
	Slicing details (direction & thickness)	67	9	2	5	21	1	1	0	3	1	110	4.09 %
	Reconstruction (software & format)	52	12	7	8	4	4	0	3	2	5	97	4.85 %
	Exper. condition	48	41	44	90	9	43	45	56	76	68	520	9.73 %
	Others ^b (incl. Completeness)	71	0	0	0	0	0	0	0	1	3	75	0.53 %
Source	Contributing lab	250	49	61	56	34	15	11	41	49	40	585	6.21 %
	Data reference	442	72	83	96	57	32	27	68	64	71	1012	6.13 %
Total		3623	334	334	349	427	185	121	421	377	1064	7250	5.06 %

7.x denotes version of the NeuroMorpho.Org. The vertical doubled line indicates the adoption of the Metadata Portal by the internal curation team

^aDevelopmental stage, Sex (continuous variables Age and Weight are not relevant here)

^bProtocol design, Objective type, Physical integrity, Structural domain, Morphological attributes (continuous variables Magnification and Tissue shrinkage are not relevant here)

Note that the growth of the data has maintained an approximately constant pace throughout the analyzed period, with similar amounts of metadata annotations considered before and after the introduction of the portal. Based on our lab records and analytics reports, the initial manual annotation of datasets in the last four releases (v.7.1–4) prior to deploying the metadata portal took an average of 1 h and 40 min per article (100 ± 10 min, mean \pm standard deviation; $N = 308$ articles). The mean time required for the same operation in the five subsequent releases following the introduction of the portal (v. 7.5–9) dropped to 55 ± 5 min per article ($N = 166$), corresponding to a net saving of 45 min in the first step of metadata curation for each dataset. Moreover, all new terms need to be identified both to ensure appropriate database updating and synchronization, and to inform users upon release. This operation used to be carried out manually by visually inspecting each form, which normally required 14 ± 1 h of labor per release. The web-based portal automatically logs and reports all new terms, thus completely eliminating the need for this effort.

After the first annotation phase, metadata curation requires a second step of quality check after the preview release on the password-protected server and corresponding review by data contributors and database curators prior to public release. In most cases, this second phase entails at least some corrections and adjustments. When metadata was entered manually through a regular spreadsheet form (through v.7.4), most errors requiring corrections consisted of spelling mistakes ('neocrotex' instead of 'neocortex') or use of non-preferred terms ('isocortex' or 'ctx'). A less common type of corrections involved conventional order of entries, as in "neocortex > medial prefrontal > right" vs. "neocortex > right > medial prefrontal". Altogether, these issues required 100 ± 15 corrections per release in the old system. Use of controlled vocabularies, dropdown menus, smart filters, and autocomplete functionality dramatically reduced these instances to as few as 15 ± 5 per release. Corrections are especially taxing on data curators and database administrators, because mistaken 'new' entries need to be removed post-ingestion to avoid inconsistencies, indices and caches cleared, and synonyms properly linked for searches to work as intended. The drastic reduction in the number of required corrections saved about 18 h of labor per release, from 22 ± 3 prior to portal adoption to 4 ± 1 afterwards.

When considering all sources of time saving (annotation, new term extraction, and corrections), the introduction of the web portal reduced the metadata annotation effort from 115.6 ± 35.4 to 48.3 ± 19.5 person-hour/release, a 58% effort reduction (Fig. A2.3).

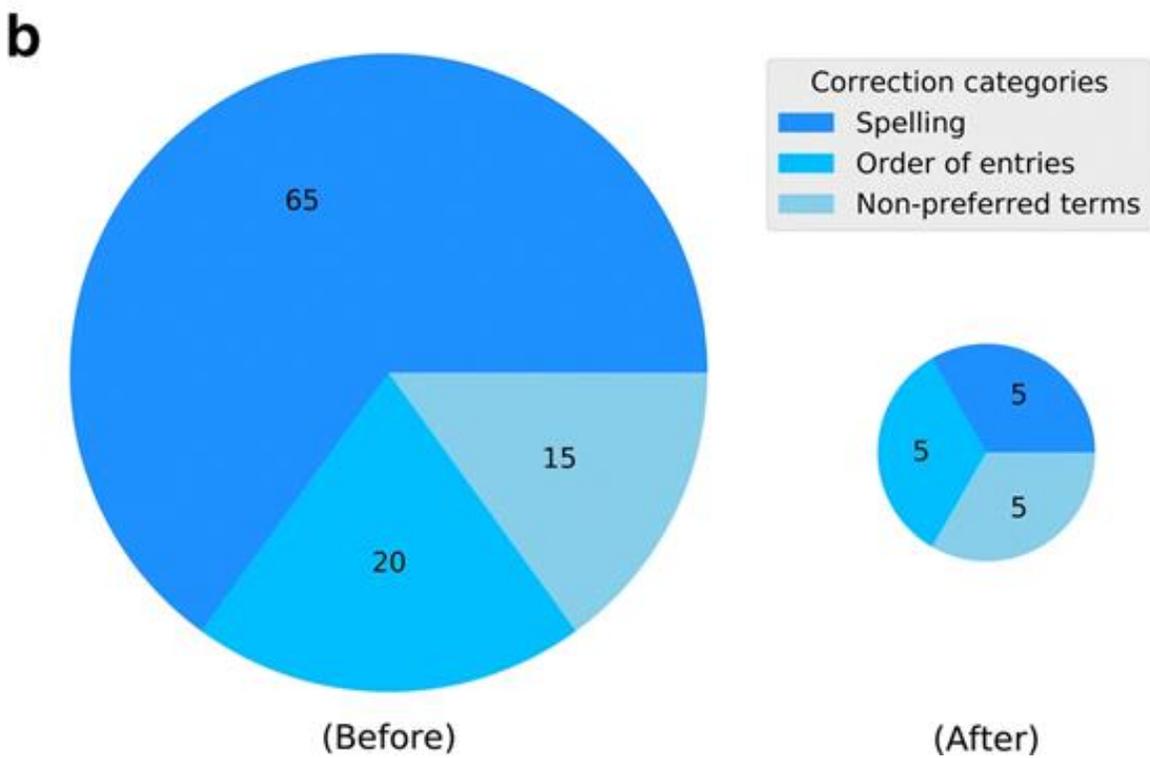
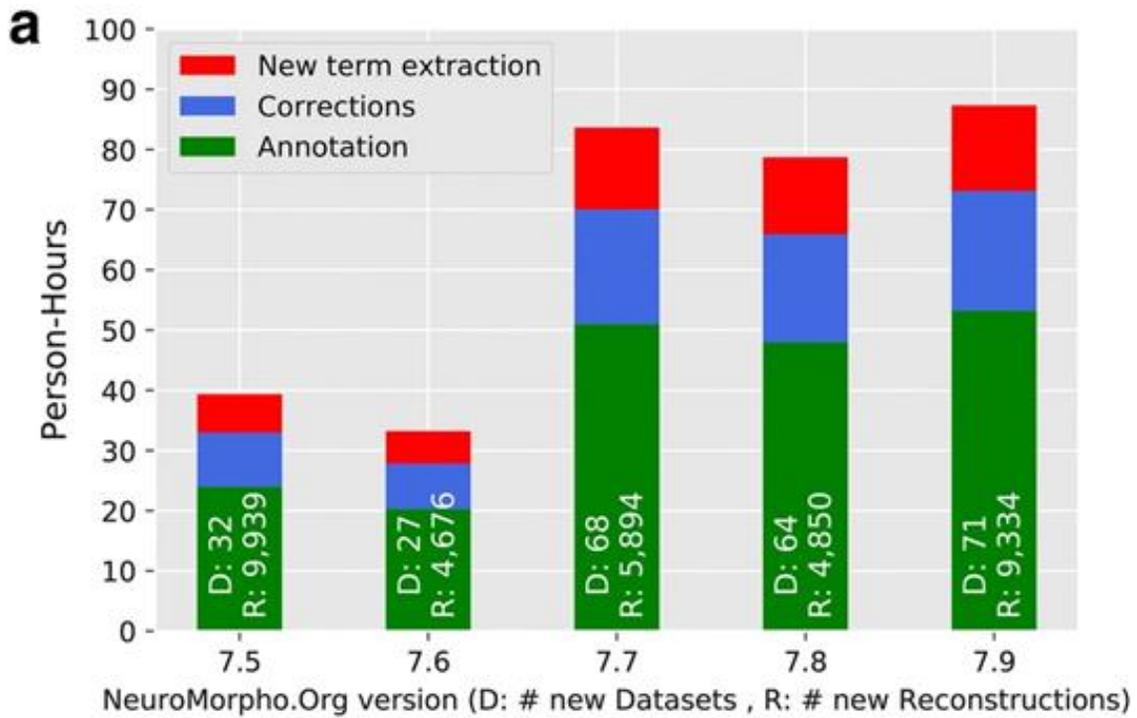


Figure A2. 3. Labor-saving per version release yielded by the metadata portal. **a** Effort saved by the adoption of web-based annotation in the last 5 releases (7.5 to 7.9) of NeuroMorpho.Org. **b** Detailed categories of annotation mistakes requiring post-ingestion corrections during the review phase before (left) and after (right) transitioning to the new annotation system; the average numbers of necessary corrections per release are indicated inside the pie charts

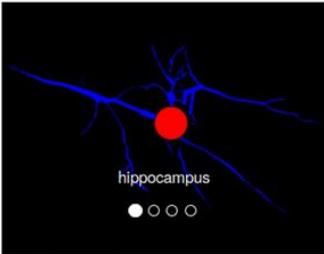
Usage protocol

In addition to the many advantages of the metadata portal described above, the web-based implementation naturally enables its direct usage by the authors of the articles described the original datasets, namely the data contributors. Considering the greatly improved performance of metadata annotation, with this article we invite all researchers depositing their neuronal and glial tracings into NeuroMorpho.Org to utilize the portal for annotating their submission. In this section, we overview the functionality, features and usage of the system <http://cng-nmo-meta.orc.gmu.edu/>.

In order to limit the server susceptibility to automated malicious activities, users must log in via username (nmo-author) and password (neuromorpho) or using a Google account. Using the latter approach, the user's entry remains private (only visible to the contributor and the administrators, but not to other users) until approved for public release by the NeuroMorpho.Org curators. Upon entering the portal (Fig. A2.4), users can create a dataset by clicking on the 'New!' button in the main view.

NMO Metadata Portal [Datasets](#) a Kayvan.Bijari

b [New!](#)



c **Datasets**

Show entries Search:

ID	Dataset	Author	Version	Status	Created	Management
2017	Sample Dataset d	John Doe	Train	Ongoing...	27/Feb/17	

Showing 1 to 1 of 1 entries First Previous **1** Next Last

© NeuroMorpho.Org

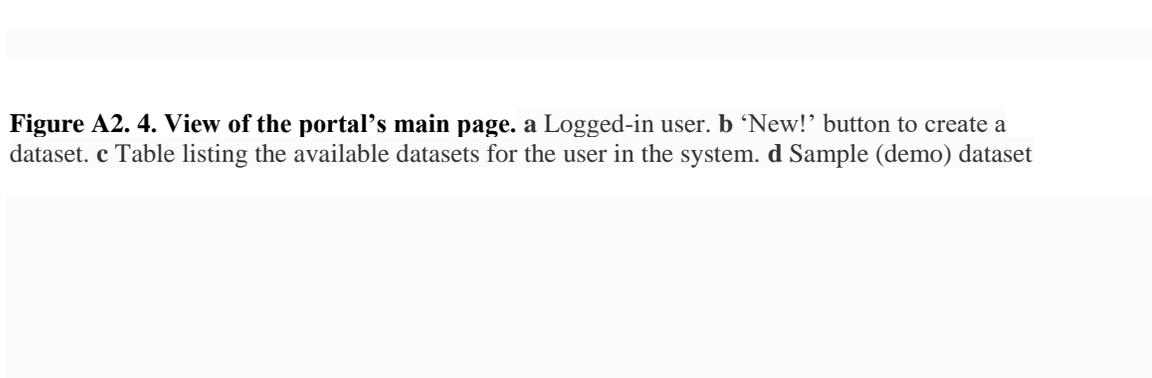


Figure A2. 4. View of the portal's main page. **a** Logged-in user. **b** 'New!' button to create a dataset. **c** Table listing the available datasets for the user in the system. **d** Sample (demo) dataset

The newly opened window prompts the insertion of information related to the reference publication such as PMID, authorship, and grant support. Next, clicking 'Submit & create the dataset' transitions to the next phase, namely uploading reconstruction files and defining the experimental groups (Fig. A2.5).

Archive: **Bijar**

NMO-version:

PMID: **135482**

Author(s): **Kayvan Bijari**

Uri: [None](#)

📌 *NIH grants R01NS39600 and U01MH114829*

added by **Kayvan.Bijari** 0 minutes ago!



a

🔔 Neural reconstruction files (groups) are not available yet!

- If you have distinct experimental conditions (e.g. anatomical location, control vs treatment, animal sex/age etc) it'd be most helpful if you separate each group of cells in different folder and zip them all together. There is an example in the [sample dataset](#) for you to follow.

- If your file size is huge (more than 1GB) please use our [FileDrop system](#) to upload your files.

Choose a file :

Browse...

No file selected.

Upload!

b

Neuron Groups: ([New](#) ✎ , [Remove](#) 🗑) [0 group(s), 0 cells]

Recently edited: Empty

No Neuron is available in this dataset yet!

c

🗨 Reviews:

No reviews yet!

Comment:

Post!

Figure A2. 5. Newly created dataset in the metadata portal. **a** Basic information about the dataset as well as edit and removes buttons. **b** Reconstruction section to upload digital tracing files. **c** Menu for creating, editing, and removing the experimental groups associated with the dataset

To upload reconstruction files, users should click the ‘Browse’ button to locate the zip folder containing the data. Separate groups with distinct experimental conditions (control vs. treatment, but also different anatomical locations, animal sex/age, etc.) must be organized as corresponding folder in the compressed archive. The ‘New’ button in the Neuron group section adds an experimental group and calls a new form window requesting the corresponding metadata details (Fig. A2.6).

NMO Metadata Portal | [Bjartur](#) | [\(new\)](#) | [Kayvan Bijani](#)

Add a

Please! fill
Hover exta

Group name
Control
Exp: Control, La

Subject
Please select...

Number of
Exp: 100
Enter the num

Species
Please select...

Gender
Please select...

Min Age
Please specify

Min Weigh
Please specify

Review Co

Anatomy
Please provide

Brain Region
Please select...
Exp: Hippocamp

Brain Region
Select Som
Exp: Granule
 Do you

Cell-type
Please provide

Cell-type S
Select Som
Exp: Type I, E
 Do you

Review Co

Experiment and Reconstruction

Experimental Pro
Please select...
Exp: In vivo, In vitro, c

Stain
Please select...
Exp: Golgi, DiI, Bloo

Slicing Direction
Please select...
Exp: transverse, hori

Objective Type
Please select...
Exp: Dry, Oil, Water

Tissue Shrinkage
Not reported
Shrinkage is reported
Shrinkage Corre
Exp: Type I, E
Shrinkage is correcte

Review Comm

Data

Numerical Units
Please select...
Micrometer or Pixel (if Pixel specify the pixel size)

Data Type
Please select...
Exp: swc, asc, nrk, dat ...

Morphological Attributes
Which of the following measurements do you consider reliable in your reconstruction files?
Dimension | **Angles** | **Diameter**
Is the Z (depth) over

Review Comments: (put

Neuron Description & Phys
Which of the following neur
slice sectioning, partial lab

Neuronal compartments
 Soma

Axon & Dendrites:

Neurites & Processes

Review Comments: (put each comment

Contributor Inform

Archive Name
Archive name:
The reconstructions
Lab / Institute / A
Please select...
Name and address

Review Comm

General Notes
Please provide any additional notes here:
Any additional notes about the reconstructions that you would like to include.

Submit!

[Save and propagate changes](#) [Save and open next group](#)

© NeuroMorpho.Org

Figure A2. 6. Metadata form to annotate the details of the reconstruction within each experimental group.

After filling out the entries as completely as possible, the user can click on ‘submit the group’. In case of multiple groups, the auxiliary buttons facilitate duplication, propagation, and modification of metadata details (Fig. A2.7).

NMO Metadata Portal Kayvan.Bijari

Archive: **Bijari** [Share] [Trash]
 NMO-version:
 PMID: **135482**
 Author(s): **Kayvan Bijari**
 Url: None
 NIH grants R01NS39600 and U01MH114829.

added by **Kayvan.Bijari** 2 minutes ago!

Groups: [Download] [Make Groups] [Remove]

click here to see groups

Choose a file :
 Choose File | No file chosen
 Upload!

Neuron Groups: ([New](#) , [Remove](#)) [3 group(s), 92 cell(s)]
 Recently edited: KO2 , KO , WT

- KO2** group with 14 cell(s) [Share] [Add] [Trash]
- KO** group with 16 cell(s) [Share] [Add] [Trash]
- WT** group with 62 cell(s) [Share] [Add] [Trash]

Reviews:
 No reviews yet!

Comment:

Post!

Figure A2. 7. Final dataset with reconstructions and all experimental groups added in the metadata portal.

Shortly after final submission, the internal NeuroMorpho.Org secondary curation begins, which includes validating the newly added terms. The reconstruction files along with the descriptive metadata are then ready for ingestion and release on a password-protected preview site that mirrors the look-and-feel of NeuroMorpho.Org while allowing

extensive review of content, annotations, and functionality by data contributors and curators prior to public release.

Conclusion

Continuous growth of neuroscience knowledge requires a parallel maturation of informatics resources to annotate data for future re-use and interpretation. This report introduced a newly developed metadata portal that leverages web-based technologies to facilitate effective curation of digital reconstructions of neuronal and glial morphologies. All components of this framework are open-source and can thus be adopted for or adapted to the needs of other related projects. Moreover, the metadata portal is ready to be integrated with artificial intelligence modules such as natural language processing or smart recommendation systems to further expedite and improve the critical bottleneck of database curation. Recently, machine learning algorithms have been successfully deployed for metadata extraction (Martínez-Romero et al., 2019). In particular, text mining tools, such as *named entity recognition*, can learn, identify, and label crucial elements of neuroscience documents like neuron names, brain regions, and experimental conditions (Bachman et al., 2018; Shardlow et al., 2019). Hence, our future aim will be, first, to train and validate a model on the growing set of curated articles in the NeuroMorpho.Org literature database, as well as on the named entities therein; and then to deploy it on the metadata portal in order to facilitate assisted keyword extraction. To

be clear, we consider it unrealistic to expect full automation of all metadata extraction tasks in the near future, as too many decisions involve domain-specific expertise and often ad-hoc conventions. Nevertheless, the prospect of a hybrid human–computer interface ergonomically optimized to maximize the breadth, depth, and accuracy of annotation while minimizing time and labor is in our view well within reach. As a first step in that direction, the systematic coding of the prior entirely manual spreadsheet annotation process of NeuroMorpho.Org metadata within a web-form interfaced to a back-end database has already substantially reduced the ongoing curation effort. We are now releasing this system publicly to allow willing data contributors to enter the details of their datasets directly at the time of data submission. While the design of the portal still allows and encourages an iterative process of collaborative review to reduce the risk of ambiguity and inconsistencies, we hope that enabling metadata annotation by the “ultimate experts” who produced the data will bring us closer to a robust, distributed, and dynamic community-based resource.

Availability of data and materials

Project name: NeuroMorpho.Org Metadata Annotation. Project home page: <http://cng-nmo-meta.org.gmu.edu/>. Operating system: Platform independent. Programming language: Python, HTML, Java script. Other requirements: Python 2.7,

Django 1.9, Nginx License: GPL 3.0. Source

code: <https://github.com/NeuroMorpho/metadata-portal>.

Acknowledgements

The authors are grateful to all the authors who have generously shared their data and filled the related metadata spreadsheets, to Dr. Bengt Ljungquist for his comments on the manuscript, and to the entire NeuroMorpho.Org team including interns for their constructive feedback on the metadata portal.

APPENDIX 3: LARGE SCALE SIMILARITY SEARCH ACROSS DIGITAL RECONSTRUCTIONS OF NEURAL MORPHOLOGY

Bengt Ljungquist¹, Masood A. Akram¹, Giorgio A. Ascoli^{1*}

¹Center for Neural Informatics, Structures, & Plasticity and Bioengineering Department, George Mason University, Mail Stop 2A1, 4400 University Dr, Fairfax, VA, United States of America

*This appendix was posted on bioRxiv (2021), doi: <https://doi.org/10.1101/2021.12.17.473026>

Abstract

Most functions of the nervous system depend on neuronal and glial morphology. Continuous advances in microscopic imaging and tracing software have provided an increasingly abundant availability of 3D reconstructions of arborizing dendrites, axons, and processes, allowing their detailed study. However, efficient, large-scale methods to rank neural morphologies by similarity to an archetype are still lacking. Using the NeuroMorpho.Org database, we present a similarity search software enabling fast morphological comparison of hundreds of thousands of neural reconstructions from any species, brain regions, cell types, and preparation protocols. We compared the performance of different morphological measurements: 1) summary morphometrics calculated by L-Measure, 2) persistence vectors, a vectorized descriptor of branching structure, 3) the combination of the two. In all cases, we also investigated the impact of applying dimensionality reduction using principal component analysis (PCA). We

assessed qualitative performance by gauging the ability to rank neurons in order of visual similarity. Moreover, we quantified information content by examining explained variance and benchmarked the ability to identify occasional duplicate reconstructions of the same specimen. The results indicate that combining summary morphometrics and persistence vectors with applied PCA provides an information rich characterization that enables efficient and precise comparison of neural morphology. The execution time scaled linearly with data set size, allowing seamless live searching through the entire NeuroMorpho.Org content in fractions of a second. We have deployed the similarity search function as an open-source online software tool both through a user-friendly graphical interface and as an API for programmatic access.

Keywords: Neuronal Morphology, Principal Component Analysis, Neuroinformatics, Similarity search, Software as a Service

Introduction

Since the dawn of neuroscience, with the elegant drawings of Ramón y Cajal (Azoulay & Ramón y Cajal, 1894), it has been known that the branching morphology of neuronal and glial (henceforth referred to as neural) arbors underlie their functional role in the nervous system. Specific neuron types often have characteristic dendritic or axonal trees, as exemplified by the iconic structures of cerebellar Purkinje cells and cortical chandelier cells, respectively (Ascoli, 2015a). Multifarious analysis approaches have

been developed both to establish common morphological characteristics within neural types and to compare structural differences between types (Laternus et al., 2020).

Quantitative representations of tree morphologies harnessed, for example, different morphometrics statistics (Y. Lu et al., 2015a; Polavaram et al., 2014c; Scorcioni et al., 2008b; Uylings & van Pelt, 2002) and spatial density maps (Cuntz, 2012; Jefferis et al., 2007). Arbor topology was captured through sequence representation (Gillette et al., 2015) or using a local alignment between pairs of neurons via dynamic programming (Wan et al., 2015c). More recently, algebraic methods based on persistent homology allowed to quantify branching distribution into vector metrics (Kanari et al., 2018b; Y. Li, Wang, et al., 2017a).

It should be possible to leverage these characterizations to find similar morphologies from a large set, given a neuronal or glial reconstruction of interest as the query vector. To this purpose, the NBLAST algorithm combined cell position with local geometry (Costa et al., 2016a). A different method for global search employed an asymmetric binary coding strategy based on the maximum inner product (Z. Li, Fang, et al., 2017), while encoding of morphology with hashing forests has shown promising performance on large datasets (Conjeti et al., 2016a). Another variant is to query sub-structures of the neuron with graph representations of the (sub-)trees (J. Yang et al., 2020) or using structure tensors and expanding this field via gradient vector flow (Ganglberger et al., 2014). A conceptually related problem is to compare two morphological reconstructions of one and the same neuron, such as when benchmarking

an automated tracing software against the gold standard of expert manual proofreading (Gillette et al., 2011).

For the past 15 years, NeuroMorpho.Org enabled sharing of 3D reconstructions of neural morphology (Akram et al., 2018b), with the latest major release, 8.1, accruing 150,000 digital tracings from 78 species, 1330 cell types, 381 brain regions, and 774 contributing labs. The growth rate of this repository has continuously increased due to a combination of more efficient reconstruction techniques, greater willingness of the neuroscience community to share, and rising open data expectations from funding organizations and scientific publishers (Ascoli et al., 2017c). In parallel, the NeuroMorpho.Org internal processing pipeline evolved into a micro-service-based architecture, reducing the time from deposition to publication from months to weeks (Anderson et al., 2021). Within the data standardization workflow, it must be verified that each submitted reconstruction is not a duplicate of any previously submitted specimen. This step, equivalent to an extreme case of similarity matching, used until recently Pearson correlation of morphometric data. However, such method was slow for larger data sets, prompting the development of the more efficient alternative described in this work.

To our knowledge, no readily available Software as a Service method provides precise and fast search of neural reconstructions solely by morphological similarity for larger data sets (>100k reconstructions). Therefore, we engineered a combination of the state-of-the-art software Facebook AI Similarity Search, or FAISS (J. Johnson et al., 2021), with L-Measure statistical morphometrics (Scorcioni et al., 2008b) and persistence

vectors descriptors (Y. Li, Wang, et al., 2017a). We report that this system enables efficient and massive parallel searches, works flexibly with both global (whole cell) and local (parts of the cell) neural arbors, and is highly effective at duplicate detection. Moreover, we investigated the impact of dimensionality reduction using PCA (principal component analysis) in terms of precision and speed. Lastly, we present the new NeuroMorpho.Org functionality enabling users to carry out similarity searches through an intuitive graphic interface or an Application Programming Interface (API). All code is released open source to ensure maximum reproducibility and encourage further community development (Gleeson et al., 2017a).

Materials and Methods

NeuroMorpho.Org provides for each reconstruction 21 morphometrics, for example, total length, average bifurcation angle, and the number of branches (for a complete list, see <http://NeuroMorpho.Org/myfaq.jsp?id=qr4#QS3>), as derived by L-Measure (Scorcioni et al., 2008b). These metrics are calculated both for the entire cell, referred to as summary morphometrics, as well as separately for each structural domain, for example only the axon, and are then referred to as detailed morphometrics. For summary morphometrics, the dimensionality is thus 21, while for detailed morphometrics it is a multiple of 21 depending on the number of distinct structural domains, up to a maximum of 63 for a neuron with apical and basal dendrites as well as an axon.

In addition to the summary and detailed morphometrics, NeuroMorpho.Org also stores 100- dimensional persistence vectors quantifying the branch distribution by

algebraic homology (Y. Li, Wang, et al., 2017a). We have modified the original C++ and Java open-source code (Y. Li, Wang, et al., 2017a) to run on Linux and added a Python wrapper to make the software callable as a service. The source code of this modified software is available open source at <https://github.com/NeuroMorpho/swc2pvec>, and the service is briefly described and callable as an API at <http://cng-nmomain.orc.gmu.edu/swp2pvec> (for a complete API description, please see the GitHub page).

NeuroMorpho.Org also provides rich metadata for all reconstructions (Parekh et al., 2015a), including both nonnumerical (e.g., species and brain region) and numerical specifications (e.g., slice thickness and objective magnification). For purposes of similarity search, we hashed and normalized nonnumerical metadata values into a numerical representation, resulting in a 29-dimensional vector.

FAISS (Facebook AI Similarity Search) is a fast and memory-efficient similarity search software developed by the company Meta, formerly Facebook (J. Johnson et al., 2021). To generate an index of distances (similarity) from the input data, FAISS offers different methods trading off properties such as memory usage vs speed. We have consistently used a flat L2 index in our implementation. The index may then be queried using a vector of the same dimensions as the indexed data. We have also normalized all data before building the index. The resulting similarity is the L2 normalized distance between the two vectors, ranging from -1 to 1, with -1 meaning they are maximally dissimilar, and 1 that they are identical. In addition to FAISS, we have implemented for benchmarking purposes a traditional Pearson correlation for similarity search of a single

cell at a time (non-parallel). We have also tested a combination of the two, FAISS similarity multiplied by Pearson correlation, which thus also can only handle searches with one neuron at a time. Pearson correlation was calculated in real-time, as pre-calculation would take for the current database 130 GB of memory, far exceeding the available RAM. The RAM footprint of a typical similarity search in our implementation stayed around 2-3 GB.

The similarity search software was written in Python, including NumPy and Flask in addition to the FAISS Python library, and deployed as a Docker service using Ubuntu Linux 18.04 LTS as the operating system on a virtual machine hosted at the data center of George Mason University's Office of Research Computing. We created the following data vectors for each neuron to build a similarity search, with dimension m as indicated:

1. Summary morphometrics ($m=21$);
2. Detailed morphometrics ($m=21, 42$ or 63 , that is, 21 per structural domain present);
3. Persistence vectors ($m=100$);
4. 1 combined with 3 ($m=121$);
5. 2 combined with 3, ($m=121, 142$ or 163);
- and 6) Binary metadata comparison ($m=29$).

For the above morphometric data sets, we applied PCA by calculating the eigenvectors (principal components) and correspondent eigenvalues of the covariance matrix and sorting them in order of falling ratio of explained variance. We adopted the broken stick method for determining how many principal components to include as base vectors (Jackson, 1993). This method advocates stopping when the ratio of explained variance of the next principal component falls below what could be expected if performing the PCA on white noise, that is $1/m$, where m is the dimension of the original

data. For the selected principal components, we calculated the sum of explained variance as a measure of their information content.

A similarity search user interface was implemented in JSP and integrated with NeuroMorpho.Org, where, starting from any cell page of interest, users may search for similar cells choosing any of the 6 methods described above. As with other NeuroMorpho.Org search functionalities, the SWC files of the found neurons or glia can be immediately visualized or saved for separate downloading. Users may also select to apply PCA and whether to utilize FAISS, Pearson correlation, or their scalar product. The graphical user interface in turn calls an underlying API, which is also independently machine-accessible (<https://github.com/NeuroMorpho/similaritysearch>).

In order to assess the capability of the software to rank morphological similarity, we performed a visual evaluation. Similarity search was performed on a representative set of 100 reconstructions from NeuroMorpho.Org selected so as to cover the broad span of species, brain regions, cell types, and experimental methods in the database. The search used summary morphometrics plus persistence vector as data, dimensionality reduction using PCA, and combined FAISS and Pearson correlation as index. The top six most similar cells were then compared with six randomly selected cells relative to the original reconstruction, and their similarity scores calculated. Furthermore, we generated a histogram of the similarity scores from each target cell to all other cells in the database.

To evaluate performance scaling for parallel searches, that is searching simultaneously through the entire NeuroMorpho.Org content for entries similar to many

cells, we used persistence vector plus summary morphometrics ($m=121$) and compared that with the same after dimensionality reduction by PCA ($m=8$). For the two approaches, we varied the number of parallel searches from 200 cells to 4800 in 200 increments and measured the similarity search execution time. This was performed using FAISS as Pearson correlation cannot handle parallel searches.

We then evaluated the ability to find data duplicates of the different morphometric indices, with and without PCA applied, thus comparing six different methods. Detailed measurements were not used as they cannot compare cells of different structural domains. Likewise, we did not use metadata similarity for the purpose of duplicate detection, as cells often share the same metadata when they are contributed from the same lab to the database. We compared the efficiency of these six methods in finding duplicates by calculating the false negative rate (FNR or miss rate) and false positive rate (FPR or fall-out) for duplicates detection. The trial for this detection utilized the 131,960 reconstructions from the latest major release of NeuroMorpho.Org that contained duplicates (v. 8.0), since those have been cleared out as of v. 8.1. Starting from a set (1000 cells) of potential duplicates identified using all three descriptors as well as an archive-by-archive inspection, we manually generated a list of 235 true duplicates by confirming through visual inspection and examination of the source files and related annotations provided by the original contributors. For all 131,960 reconstructions, we then applied parallel FAISS similarity search for an all-against-all duplicate detection using all six methods separately. This generated a list of the 10 most similar cells for each method and cell. If any of these had a similarity score ≥ 0.9999 , it was considered a

potential duplicate. This was then compared against the list of true duplicates and, if confirmed, considered a true positive, otherwise a false positive. If a cell pair was on the list as a true duplicate, but fell below the similarity threshold, it was considered a false negative. Lastly, cell pairs that neither were on the true duplicate list nor passed the similarity threshold were considered true negatives. We also evaluated impact to FPR and FNR when setting the required similarity score to ≥ 0.99999 or ≥ 0.999 .

The implementation is provided open source at the NeuroMorpho.Org GitHub account: <https://github.com/NeuroMorpho/similarity-search>. Database credentials have been removed, as direct database access is not provided to outside users for security reasons, but all data used for similarity search is provided as Python pickle files (*.pkl). It is also possible to download the application as a Docker container image, including the pickle files: <https://hub.docker.com/repository/docker/neuromorpho/sis>.

Results

The visual evaluation of representative cells demonstrated that the similarity search selects visually similar neurons when compared to random neurons (Figure A3.1). All other inspected cells showed analogous results in terms of the ability to find visually similar cells. Moreover, we noted that the distribution of similarity scores for an individual cell against the whole database was shaped differently (unimodal or bimodal, right-tailed or left-tailed etc.) for distinct cell types, such as projection neurons, interneurons, and glia.

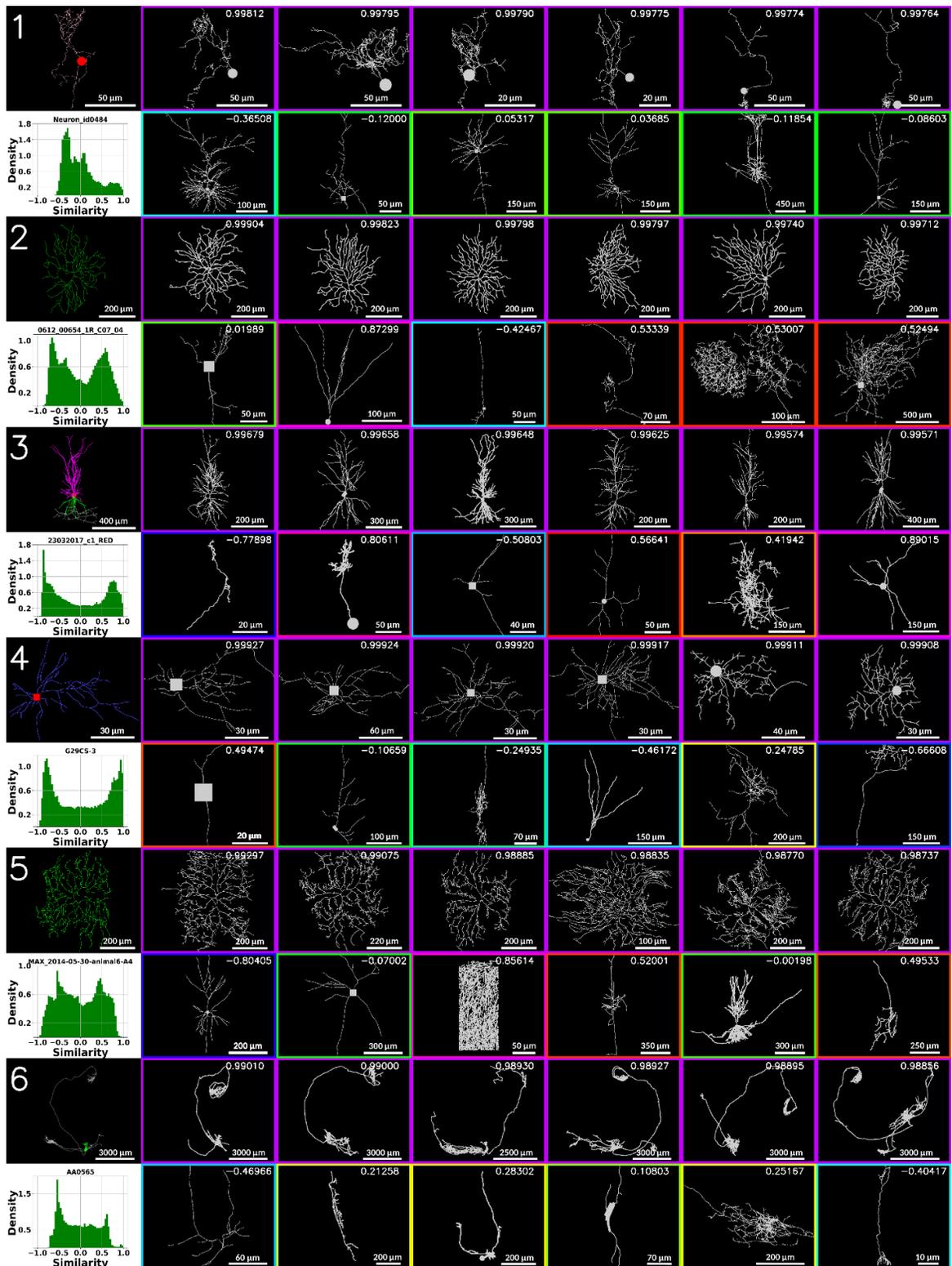


Figure A3. 1. Similarity search results compared to random controls for six representative cells from NeuroMorpho.Org: 1) zebrafish main olfactory bulb interneuron (Meehan et al., 2011) (NeuroMorpho identifier: NMO_93878); 2) mouse retinal ganglion cell (Favero & Castro-Alamancos, 2013) (NMO_128003); 3) mouse hippocampal pyramidal cell (Banks et al., 2011) (NMO_94734); 4) rabbit corpus callosum astrocyte (Silberberg & Markram, 2007) (NMO_53207); 5) drosophila melanogaster peripheral nervous system sensory neuron (Mott et al., 1997) (NMO_76528); and 6) mouse hippocampal projection cell (Amatrudo et al., 2012b) (NMO_121747). Top left figure in each row is the search target. The histogram directly below each target cell plots the similarity of that cell against all others in the database. Immediately to the right of each target cell are the six most similar cells found by the search in falling order of similarity score (indicated in top right corner of each cell). On the second row, to the right of the histogram, six random cells are shown with their similarity score in the top right corner as well. The color of the frame of each cell designates its similarity, with purple the most similar (score 1), green neutral (score 0), and blue the least similar (score -1). The original target cells are colored with respect to structural domain, while all other cells are greyed out to avoid visual bias.

The implemented microservice-based similarity search allows for fast and efficient search both for machine and human users. In particular, starting from any “cell page” (e.g., <http://neuromorpho.org/byRandom.jsp>), NeuroMorpho.Org users can search for similar morphologies after selecting preferred search parameters (Figure A3.2). Choices include the similarity implementation to use (FAISS and/or Pearson correlation), the numerical descriptors (summary morphometrics, persistence vectors, their combination, detailed morphometrics, or metadata), whether or not to apply PCA, and the result size (10, 25, 50, or 100 most similar hits). Results are delivered within a fraction of a second in all cases. The underlying API has the same parameter options for programmatic access.



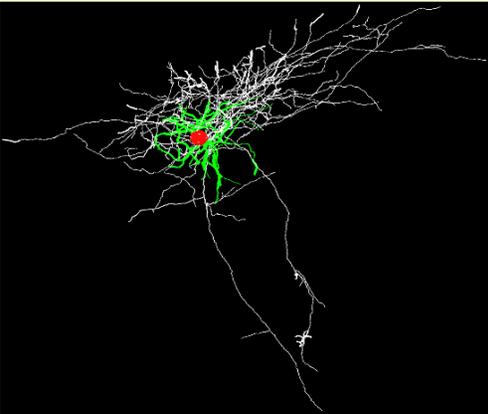
NeuroMorpho.Org

Version 8.1.37 – Released: 2021-08-12 – Content: 153622 neurons



HOME
BROWSE
SEARCH
LITERATURE COVERAGE
TERMS OF USE
HELP


[Search](#) > [Metadata](#) > [Cell Type](#) > [Interneuron](#)



Details about selected cell

NeuroMorpho.Org ID : NMO_110395

Archive Name : Kawaguchi

Species Name : Rat

Primary Brain Region : Neocortex

Primary Cell Class : Interneuron

Experimental Condition : Control

Staining Method : Biocytin

Reconstruction Method : Neurolucida

...

[Morphology File \(Standardized\)](#)
[Morphology File \(Original\)](#)
[Log File \(Standardized\)](#)
[Log File \(Original\)](#)

Get above files zipped

[3D Cell Viewer – Java, legacy](#)
[3D Cell Viewer – WebGL, novel](#)
[Animation](#)

Similar Neurons

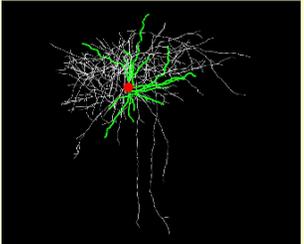
Summary + Persistence	Cells
Detailed + Persistence	10 ▼
Persistence	25
Summary	50
Detailed	100
Metadata	100

PCA No PCA

Search Method

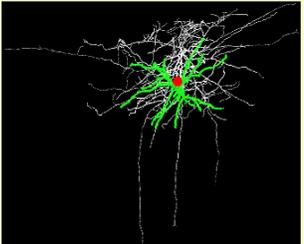
FAISS Correlation
 Correlation + FAISS

Tolias>Interneuron



Similarity: 0.9168085

Staiger>Interneuron



Similarity: 0.9024871

23 more...

Figure A3. 2. Similarity search with the graphical user interface. Starting from any cell page as in the illustrated example of an interneuron (Kubota, 2014) (left), users can access the new Similarity Search functionality and select the desired options (center, bottom). The results are displayed in order of similarity scores among the entire content of the database (right). The illustration has been adapted slightly from the current user interface look for optimal display

Search response time was linear with the number of parallel searches: 0.15s for 1000 cells, 0.3s for 2000, and 0.6s for 4000. The time difference between summary morphometrics plus persistence vectors with PCA (9 dimensions) and no PCA (121 dimensions) applied was constant with a mean of 0.05s.

The broken stick method for selecting principal components of the three descriptors resulted in different results for each (Figure A3.3). For the persistence vectors, summary morphometrics, and their combination, the number of eigenvalues greater than the cut-off limit were 7, 4, and 8, with a sum of explained variance of 96.5%, 91.3%, and 97.1% respectively.

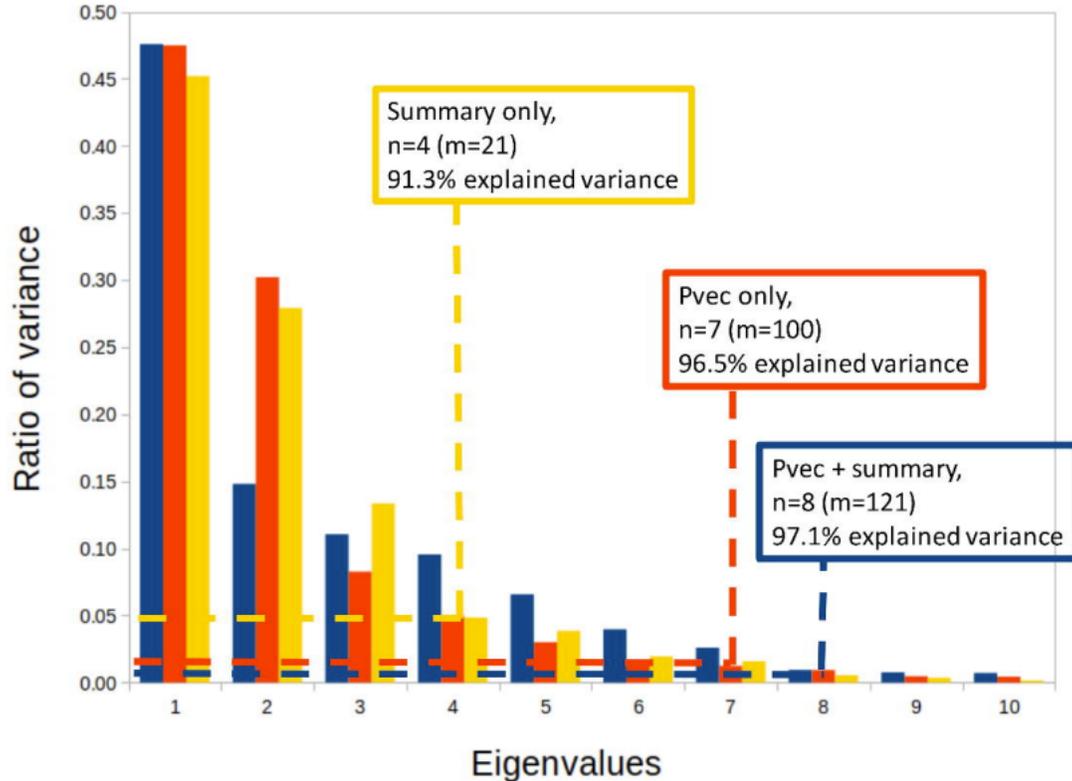


Figure A3. 3. Ratio of variance of each eigenvalue for the three different descriptors used in similarity search. Dashed lines represent the calculated cut-off n using the broken stick method ($>1/m$ where m is the original dimension).

Evaluation of the duplicate detection test showed that combining summary morphometrics and persistence vectors at a similarity threshold of 0.9999 is effective for finding the occasional repeated reconstructions (Table A3.1). Notably, persistence vectors alone performed worst, while summary morphometrics performed better but still far worse than the combined method both in terms of false positive rate (FPR) and false negative rates (FNR). Applying PCA improved both FPR and FNR for persistence vectors, while performing worse for FNR using summary morphometrics and slightly

worse for the combined method, but substantially better for FPR for both. Using a higher similarity threshold of 0.99999 improved FPR slightly to 0.18% but worsened FNR dramatically to 29.87% for the combined method with PCA. Conversely, reducing the threshold to 0.999 completely eliminated false negative but yielded an unacceptably high FPR of 7.4%.

Table A3. 1. Duplicate detection false positive rate (FPR) and false negative rate (FNR) for the different methods at a threshold similarity of 0.9999, expressed as percentages with standard deviation. A total of 235 true duplicates were presented in the set of 131,960 cells used for the statistics calculation

Method	Full - FPR	PCA - FPR	Full - FNR	PCA - FNR
Persistence vectors	35.37% ± 0.13%	19.91% ± 0.11%	24.68% ± 0.12%	17.02% ± 0.10%
Summary Morphometrics	17.34% ± 0.10%	6.66% ± 0.07%	2.55% ± 0.04%	9.79% ± 0.08%
Persistence vectors + summary morphometrics	1.34% ± 0.03%	0.49% ± 0.02%	0.43% ± 0.02%	0.85% ± 0.03%

Discussion

In this work, we introduced and evaluated a similarity search among neuronal and glial reconstructions based on three numerical descriptors of morphology. When combined with the Facebook AI Similarity Search (FAISS) software, the result is fast, efficient, and precise. In addition to sharing the code freely for further development and standalone implementations, we deployed this new function in the publicly available database NeuroMorpho.Org both as a user-friendly graphical interface and as API. Users can perform seamless similarity searches against hundreds of thousands of cells easily and quickly. Common applications include finding reconstructions similar to a target archetype of interest or identifying occasional duplicates across large data sets. The results obtained with FAISS are qualitatively similar to those yielded by traditional Pearson correlation, but FAISS additionally enables massive parallel searches. This allows the programmatic ranking of similarities for thousands of reconstructions simultaneously with results in fractions of a second. Such a performance, which to our knowledge has not been previously achieved, may facilitate ever more powerful unbiased classification of neural morphology (Bijari et al., 2021a).

A visual inspection of the similarity results showed that the method has a strong ability to find similar reconstructions to a representative variety of target morphologies when compared to a random selection. Interestingly, the distributions of the similarity

scores against the entire NeuroMorpho.Org database varied substantially from one target cell to another. We observed that cells with similar metadata tended to display comparable similarity distributions, but a detailed investigation of these distributions remains a subject for future study.

We evaluated the information content of the three morphological descriptors: persistence vectors, L-Measure summary morphometrics, and the combination of the two, by studying the explained variance after PCA application. Persistence vectors combined with summary morphometrics constitute the most information rich descriptor, with a reduction from 121 to 8 dimensions retaining over 97% of the variance. The summary morphometrics, with PCA dimensionality reduction from 21 to 4, only retained 91.3% of the variance, while the persistence vectors, with dimensionality reduced from 100 to 7 post-PCA, retained 96.5% of the original variance. Our interpretation is that summary morphometrics alone is a noisier descriptor but delivers superior precision for similarity search when combined with persistence vectors.

Execution time performance scales linearly with a growing number of cells in a parallel query. However, larger dimensions of each query vector have a comparatively smaller impact, as the similarity with no PCA applied ($m=121$) was only marginally slower than the similarity search with PCA applied ($m=8$). It should be noted that we have used the CPU implementation of FAISS on standard server hardware, with no GPU (Graphics Processing Unit) support. This was sufficient for our purposes, as the similarity search in most cases returned results within a fraction of a second. The search is expected to execute substantially faster when applying GPU acceleration. The method is therefore

likely suitable also for exceptionally large data sets including whole-brain data sets of millions of neural reconstructions.

Using persistence vectors and summary morphometrics combined as a descriptor constitutes in our experience the strongest similarity search function. This same combination also provides the most effective duplicate detection test and is superior to the persistence vectors and summary morphometrics used separately as query vectors. We also noticed that applying PCA generally yields improved results for all three methods investigated when considering both false positive and negative rates. Therefore, we recommend the combined method with PCA as the default setting for NeuroMorpho.Org similarity searches, although different options should be explored when working with defined subgroups of cells, such as only microglia, only cultured cells, only long axonal projections, or only invertebrate neurons.

Similarity search is available for both human users through the NeuroMorpho.Org graphical interface as well as for programmatic usage as an API, therefore allowing all users to discover morphologically similar neurons efficiently. The now fully automated microservice-based data ingestion pipeline of NeuroMorpho.Org (Anderson et al., 2021) checks for duplicate reconstructions before ingestion as a part of the data quality assurance. This is necessary as reconstructions may unintentionally be resubmitted if they are used in different studies together with novel reconstructions. Lastly, this work highlights NeuroMorpho.Org as a mature database, which provides the necessary data diversity and quality required for a precise similarity search.

Acknowledgements

The authors thank Praveen Menon, Navy Merianda, and Pruthvik Narayanaswamy for their contributions in development of a prototype for similarity search. This research was supported in part by NIH grants U01MH114829, R01NS39600, and R01NS08608

APPENDIX 4: SCALE CORRECTION

Glia

1. Zheng

PMID 30415998

Scale bar in the article = 5 μm

Pixel for scale bar = 33 pixel

Height of the cell in pixel = 186 pixel

Nominal height reported = 249.47 μm

33 is 5.63 times of 186, so $5 \mu\text{m} \times 5.63 = 28.1 \mu\text{m}$ (this is the Height)
 $28.1/249.47=0.1130$, ABEL should be multiplied by 0.1130

2. Foerster

PMID 29162696

Scale bar in the article = 10 μm

Pixel for scale bar = 37 pixel

Height of the cell in pixel = 367 pixel

Nominal height reported = 345.58 μm

37 is 9.91 times of 367, so $10 \mu\text{m} \times 9.91 = 99.1 \mu\text{m}$ (this is the Height)
 $99.1/345.58= 0.2870$, ABEL should be multiplied by 0.2870

3. Di Benedetto

PMID 26869881

Scale bar in the article = 20 μm

Pixel for scale bar = 34 pixel

Height of the cell in pixel = 50 pixel

Nominal height reported = 222.35 μm

34 is 1.47 times of 50, so $20 \mu\text{m} \times 1.65625 = 29.411 \mu\text{m}$ (this is the Height)
 $29.411/222.35= 0.1323$, ABEL should be multiplied by 0.1323

4. Wake

PMID 31862977

Scale bar in the article = 10 μm

Pixel for scale bar = 20 pixel

Height of the cell in pixel = 115 pixel

Nominal height reported = 360.78 μm

20 is 5.75 times of 115, so $10 \mu\text{m} \times 5.75 = 57.5 \mu\text{m}$ (this is the Height)

$57.5/360.78 = 0.1594$, ABEL should be multiplied by 0.1594

5. Maguire-Zeiss

PMID 28921719

Scale bar in the article = 10 μm

Pixel for scale bar = 37 pixel

Height of the cell in pixel = 89 pixel

Nominal height reported = 339.08 μm

37 is 2.405 times of 89, so $10 \mu\text{m} \times 2.405 = 24.05 \mu\text{m}$ (this is the Height)

$24.05/339.08 = 0.0710$, ABEL should be multiplied by 0.0710

6. Rusakov

PMID 30177844

Scale bar in the article = 10 μm

Pixel for scale bar = 24 pixel

Height of the cell in pixel = 201 pixel

Nominal height reported = 358.16 μm

24 is 8.375 times of 201, so $10 \mu\text{m} \times 8.375 = 83.75 \mu\text{m}$ (this is the Height)

$83.75/358.16 = 0.2338$, ABEL should be multiplied by 0.2338

7. H_Zhang

PMID 30654821

Scale bar in the article = 20 μm

Pixel for scale bar = 38 pixel

Height of the cell in pixel = 128 pixel

Nominal height reported = 334.26 μm

38 is 3.36 times of 128, so $20 \mu\text{m} \times 3.36 = 67.36 \mu\text{m}$ (this is the Height)

$67.36/334.26 = 0.2015$, ABEL should be multiplied by 0.2015

8. Weil

PMID 26833850

Scale bar in the article = 20 μm

Pixel for scale bar = 78 pixel

Height of the cell in pixel = 81.9 pixel

Nominal height reported = 30.57 μm

78 is 1.05 times of 81.9, so $20 \mu\text{m} \times 1.05 = 21.17 \mu\text{m}$ (this is the Height)

$21.17/30.57 = 0.6926$, ABEL should be multiplied by 0.6926

9. Fernandez-Ruiz

PMID 30076846

Multiply by 0.65 & Divide by 2

10. Xiong

PMID 34330901

Scale bar in the article = 20 μm

Pixel for scale bar = 124 pixel

Height of the cell in pixel = 302 pixel

Nominal height reported = 237.32 μm

124 is 2.43 times of 302, so $20 \mu\text{m} \times 2.43 = 48.6 \mu\text{m}$ (this is the Height)

$48.6/237.32 = 0.2050$, ABEL should be multiplied by 0.2050

Neurons

1. Firestein

PMID 25542305

Scale bar in the article 50 μm

Pixel of scale bar = 76 pixel

Height of the cell in pixel = 323 pixel

Nominal height reported = 474.15 μm

76 is 4.25 times of 323, so $50 \mu\text{m} \times 4.25 = 212.5 \mu\text{m}$ (this is the Height)

$212.5/474.15 = 0.4482$, ABEL should be multiplied by 0.4482

PMID 29982499

Scale bar in the article 50 μm

Pixel of scale bar = 70 pixel

Height of the cell in pixel = 273 pixel

Nominal height reported = 27.24 μm

70 is 3.9 times of 273, so $50 \mu\text{m} \times 3.9 = 195 \mu\text{m}$ (this is the Height)

$195/27.24 = 7.1586$, ABEL should be multiplied by 7.1586

PMID 32157575

Scale bar in the article 100 μm

Pixel of scale bar = 83 pixel

Height of the cell in pixel = 188 pixel

Nominal height reported = 28.79 μm

83 is 2.265 times of 188, so $100 \mu\text{m} \times 3.34 = 226.5 \mu\text{m}$ (this is the Height)

$226.5/28.79 = 7.8675$, ABEL should be multiplied by 7.8675

2. Moons

PMID 34073191

Scale bar in the article 20 μm

Pixel of scale bar = 36 pixel

Height of the cell in pixel = 399 pixel

Nominal height reported = 174.54 μm

36 is 11.08 times of 399, so $20 \mu\text{m} \times 11.09 = 221.8 \mu\text{m}$ (this is the Height)

$221.8/174.54=1.2710$, ABEL should be multiplied by 1.2710

3. Wong_Silver

PMID 30074985

Scale bar in the article 50 μm

Pixel of scale bar = 68 pixel

Height of the cell in pixel = 423 pixel

Nominal height reported = 208.5 μm

68 is 6.22 times of 423, so $50 \mu\text{m} \times 6.22 = 311 \mu\text{m}$ (this is the Height)

$311/208.5= 1.4920$, ABEL and Height should be multiplied by 1.4920

4. Manica_Leon

PMID 32633719

Scale bar in the article 50 μm
Pixel of scale bar = 71 pixel
Height of the cell in pixel = 475 pixel
Nominal height reported = 124.82 μm

71 is 6.69 times of 475, so $50 \mu\text{m} \times 6.69 = 334.5 \mu\text{m}$ (this is the Height)
 $334/124.82 = 2.6760$, ABEL and Height should be multiplied by 2.6760

5. Wadiche

PMID 21490706

Scale bar in the article 20 μm
Pixel of scale bar = 72 pixel
Height of the cell in pixel = 420 pixel
Nominal height reported = 96.19 μm

72 is 5.83 times of 420, so $20 \mu\text{m} \times 5.83 = 116 \mu\text{m}$ (this is the Height)
 $111/96.19 = 1.1540$, ABEL and Height should be multiplied by 1.1540

6. Cai

PMID 30715234

Scale bar in the article 50 μm
Pixel of scale bar = 15 pixel
Height of the cell in pixel = 83 pixel
Nominal height reported = 95.24 μm

15 is 5.53 times of 83, so $50 \mu\text{m} \times 5.53 = 276.5 \mu\text{m}$ (this is the Height)
 $276.5/95.24 = 2.9032$, ABEL and Height should be multiplied by 2.9032

7. Summavielle

PMID 28274785

Scale bar in the article 20 μm
Pixel of scale bar = 22 pixel
Height of the cell in pixel = 233 pixel
Nominal height reported = 37.04 μm

22 is 10.59 times of 233, so $20 \mu\text{m} \times 10.59 = 211.8 \mu\text{m}$ (this is the Height)
 $211.8/37.04 = 5.7120$, ABEL and Height should be multiplied by 5.7120

8. Cox

PMID: 21811639

Scale bar in the article = 50 μm
Pixel for scale bar = 118 pixel
Height of the cell in pixel = 309 pixel
Nominal height reported = 473.95 μm

21 is 14.71 times of 309, so $50 \mu\text{m} \times 14.71 = 735.71 \mu\text{m}$ (this is the Height)
 $735.71/473.95 = 1.5523$, ABEL should be multiplied by 1.5523

PMID: 30395636

Scale bar in the article = 100 μm
Pixel for scale bar = 47 pixel
Height of the cell in pixel = 292 pixel
Nominal height reported = 588.73 μm

47 is 6.21 times of 292, so $100 \mu\text{m} \times 6.21 = 621 \mu\text{m}$ (this is the Height)
 $621/588.73 = 1.0553$, ABEL should be multiplied by 1.0553

9. Zhang_X

PMID 28263300

Scale bar in the article = 100 μm
Pixel for scale bar = 57 pixel
Height of the cell in pixel = 259 pixel
Nominal height reported = 92.17 μm

57 is 4.54 times of 259, so $100 \mu\text{m} \times 4.54 = 454.38 \mu\text{m}$ (this is the Height)
 $454.38/92.17 = 4.9299$ ABEL should be multiplied by 4.9299

CONCLUSION

As a doctoral student, I made substantial contributions to NeuroMorpho.Org. By the time I joined the project, NeuroMorpho.Org hosted 62,000 reconstructions, and with the recent version of the database, the total content has exceeded 172,000 digital tracings. I was specifically responsible for adding glia to the database in 2017 with an initial number of 195, and with the recent release, these numbers total more than 19,000. As the amount of data increases, there is always a need to add more functionalities to NeuroMorpho.Org. I wanted to organize the data in a way that was conducive to datamining, exploratory analysis, and machine learning. Therefore, we added summary reporting to the database, which allows users to download the data of their choice quickly and efficiently.

I leveraged summary reporting to download a balanced dataset of glia and neurons for supervised classification. All classifiers performed exceptionally well in separating glia and neurons. Previous evidence suggests that neuronal dendrites are bigger than glial processes, therefore upon investigation of morphometric features contributing to this separation, I discovered a novel morphometric biomarker called average branch Euclidean length (ABEL) that separates neurons and glia better than height. We also found out that the reconstruction process for classification can be avoided by just taking ABEL of 5 branches with any image processing software with a classification accuracy of more than 95%.

The successful completion of this thesis provides new avenues for cellular comparison in the nervous system. Despite our research demonstrating the ability to distinguish glia and neurons by their arbors, the question of identifying the main subclasses within glia and neurons still remains. Traditionally, classification techniques in neuroscience have focused on neurons (Bota & Swanson, 2007a; Zhang et al., 2021b), it is exciting to see if we can differentiate between the main subclasses of neurons and glia simultaneously. We decided to perform multiclass classification using SVM, KNN, and RF. While SVM performed the best with 94.70% accuracy, majority of the misclassified cells were astrocytes and oligodendrocytes indicating that these results are still unsuitable for scientific interpretation. Such an endeavor would require balanced datasets where the paucity of certain classes (astrocytes, oligodendrocytes) does not negatively impact performance of classification algorithms.

In order to conduct a comprehensive morphometric analysis, cellular reconstructions are required. The process of reconstructing a brain cell is labor-intensive and time consuming. The fact that ABEL alone performed well especially when randomly selecting a few branches for classification, it is intriguing to see if it could be used in the future studies as a quick tool for classification of the nervous system cells. In case it does not perform well, we will still need to reconstruct a whole cell and perform morphometric analysis using supervised machine learning.

There are various neurodegenerative disorders in which the structure of glial cells is altered. For example, enlargement of microglia in Alzheimer's disease (Plaza-Zabala et al., 2017), dysfunction of oligodendrocytes leading to demyelination in multiple sclerosis (Dulamea, 2017), and changes in astrocyte morphology in Parkinson's and Huntington's disease (Liddel et al., 2017). It is fascinating to see that the same machine learning techniques that we applied in this thesis could be applied to diseased states and control conditions for the identification of abnormal cells.

There are also some limitations when it comes to classification of the nervous system cells in order to complete the cell census of the brain of any specie. The first thing is the lack of data from different species, anatomical regions, cell classes, experimental conditions, etc. Secondly, scientists have been trying to classify neurons for centuries, but glia have given a lot less attention, which is also clear from the content of NeuroMorpho.Org. Additionally, an average time to reconstruct a brain cell takes about 6 hours after all the advancements in microscopy and tracing techniques. For example, if we want to reconstruct each neuron and glia in the human brain, hundreds and thousands of hours are needed to reconstruct every cell. As a result, it will take years to study just one human brain, knowing the fact that each brain is different and works differently out of billions of other brains in the whole world. Although the field of neuroscience has made tremendous advances over the years, we still need new

techniques that will allow us to study the brain efficiently, especially if we are to complete the cell census of the nervous system.

REFERENCES

- Abdi, H., & Williams, L. J. (2010). Principal component analysis: Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Agrawal, A., & Rasouli, M. (2019). *EndNote 1-2-3 Easy!: Reference Management for the Professional*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-24889-5>
- Aguzzi, A., Barres, B. A., & Bennett, M. L. (2013a). Microglia: Scapegoat, Saboteur, or Something Else? *Science*, 339(6116), 156–161. <https://doi.org/10.1126/science.1227901>
- Aguzzi, A., Barres, B. A., & Bennett, M. L. (2013b). Microglia: Scapegoat, saboteur, or something else? *Science (New York, N.Y.)*, 339(6116), 156–161. <https://doi.org/10.1126/science.1227901>
- Aha, D. W., Kibler, D., & Albert, M. K. (1991a). Instance-based learning algorithms. *Machine Learning*, 6(1), 37–66. <https://doi.org/10.1007/BF00153759>
- Aha, D. W., Kibler, D., & Albert, M. K. (1991b). Instance-based learning algorithms. *Machine Learning*, 6(1), 37–66. <https://doi.org/10.1007/BF00153759>
- Ahmed, S., Gull, A., Khuroo, T., Aqil, Mohd., & Sultana, Y. (2017). Glial Cell: A Potential Target for Cellular and Drug Based Therapy in Various CNS Diseases. *Current Pharmaceutical Design*, 23(16). <https://doi.org/10.2174/1381612823666170316124500>

- Ahuja, C. S., Wilson, J. R., Nori, S., Kotter, M. R. N., Druschel, C., Curt, A., & Fehlings, M. G. (2017). Traumatic spinal cord injury. *Nature Reviews Disease Primers*, 3(1), 17018. <https://doi.org/10.1038/nrdp.2017.18>
- Akins, V. T., Weragalaarachchi, K., Picardo, M. C. D., Revill, A. L., & Del Negro, C. A. (2017). Morphology of Dbx1 respiratory neurons in the preBötzing complex and reticular formation of neonatal mice. *Scientific Data*, 4(1), 170097. <https://doi.org/10.1038/sdata.2017.97>
- Akram, M. A., Ljungquist, B., & Ascoli, G. A. (2022a). Efficient metadata mining of web-accessible neural morphologies. *Progress in Biophysics and Molecular Biology*, 168, 94–102. <https://doi.org/10.1016/j.pbiomolbio.2021.05.005>
- Akram, M. A., Ljungquist, B., & Ascoli, G. A. (2022b). Efficient metadata mining of web-accessible neural morphologies. *Progress in Biophysics and Molecular Biology*, 168, 94–102. <https://doi.org/10.1016/j.pbiomolbio.2021.05.005>
- Akram, M. A., Nanda, S., Maraver, P., Armañanzas, R., & Ascoli, G. A. (2018a). An open repository for single-cell reconstructions of the brain forest. *Scientific Data*, 5, 180006. <https://doi.org/10.1038/sdata.2018.6>
- Akram, M. A., Nanda, S., Maraver, P., Armañanzas, R., & Ascoli, G. A. (2018b). An open repository for single-cell reconstructions of the brain forest. *Scientific Data*, 5(1), 180006. <https://doi.org/10.1038/sdata.2018.6>
- Akram, M. A., Nanda, S., Maraver, P., Armañanzas, R., & Ascoli, G. A. (2018c). An open repository for single-cell reconstructions of the brain forest. *Scientific Data*, 5(1), 180006. <https://doi.org/10.1038/sdata.2018.6>

Amatrudo, J. M., Weaver, C. M., Crimins, J. L., Hof, P. R., Rosene, D. L., & Luebke, J.

I. (2012a). Influence of highly distinctive structural properties on the excitability of pyramidal neurons in monkey visual and prefrontal cortices. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(40), 13644–13660. <https://doi.org/10.1523/JNEUROSCI.2581-12.2012>

Amatrudo, J. M., Weaver, C. M., Crimins, J. L., Hof, P. R., Rosene, D. L., & Luebke, J.

I. (2012b). Influence of Highly Distinctive Structural Properties on the Excitability of Pyramidal Neurons in Monkey Visual and Prefrontal Cortices. *Journal of Neuroscience*, *32*(40), 13644–13660. <https://doi.org/10.1523/JNEUROSCI.2581-12.2012>

Anderson, K. R., Harris, J. A., Ng, L., Prins, P., Memar, S., Ljungquist, B., Fürth, D.,

Williams, R. W., Ascoli, G. A., & Dumitriu, D. (2021). Highlights from the Era of Open Source Web-Based Tools. *The Journal of Neuroscience*, *41*(5), 927–936. <https://doi.org/10.1523/JNEUROSCI.1657-20.2020>

Armañanzas, R., & Ascoli, G. A. (2015a). Towards the automatic classification of neurons. *Trends in Neurosciences*, *38*(5), 307–318.

<https://doi.org/10.1016/j.tins.2015.02.004>

Armañanzas, R., & Ascoli, G. A. (2015b). Towards the automatic classification of neurons. *Trends in Neurosciences*, *38*(5), 307–318.

<https://doi.org/10.1016/j.tins.2015.02.004>

- Armañanzas, R., & Ascoli, G. A. (2015c). Towards the automatic classification of neurons. *Trends in Neurosciences*, 38(5), 307–318.
<https://doi.org/10.1016/j.tins.2015.02.004>
- Ascoli, G. A. (2006). Mobilizing the base of neuroscience data: The case of neuronal morphologies. *Nature Reviews Neuroscience*, 7(4), 318–324.
<https://doi.org/10.1038/nrn1885>
- Ascoli, G. A. (2007). Successes and Rewards in Sharing Digital Reconstructions of Neuronal Morphology. *Neuroinformatics*, 5(3), 154–160.
<https://doi.org/10.1007/s12021-007-0010-7>
- Ascoli, G. A. (2015a). *Trees of the Brain, Roots of the Mind*.
- Ascoli, G. A. (2015b). Sharing Neuron Data: Carrots, Sticks, and Digital Records. *PLoS Biology*, 13(10), e1002275. <https://doi.org/10.1371/journal.pbio.1002275>
- Ascoli, G. A. (2015c). Sharing Neuron Data: Carrots, Sticks, and Digital Records. *PLOS Biology*, 13(10), e1002275. <https://doi.org/10.1371/journal.pbio.1002275>
- Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007a). NeuroMorpho.Org: A Central Resource for Neuronal Morphologies. *Journal of Neuroscience*, 27(35), 9247–9251. <https://doi.org/10.1523/JNEUROSCI.2055-07.2007>
- Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007b). NeuroMorpho.Org: A Central Resource for Neuronal Morphologies. *Journal of Neuroscience*, 27(35), 9247–9251. <https://doi.org/10.1523/JNEUROSCI.2055-07.2007>

- Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S., & Armañanzas, R. (2017a). Win-win data sharing in neuroscience. *Nature Methods*, *14*(2), 112–116.
<https://doi.org/10.1038/nmeth.4152>
- Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S., & Armañanzas, R. (2017b). Win-win data sharing in neuroscience. *Nature Methods*, *14*(2), 112–116.
<https://doi.org/10.1038/nmeth.4152>
- Ascoli, G. A., Maraver, P., Nanda, S., Polavaram, S., & Armañanzas, R. (2017c). Win-win data sharing in neuroscience. *Nature Methods*, *14*(2), 112–116.
<https://doi.org/10.1038/nmeth.4152>
- Ascoli, G. A., & Wheeler, D. W. (2016). In search of a periodic table of the neurons: Axonal-dendritic circuitry as the organizing principle: Patterns of axons and dendrites within distinct anatomical parcels provide the blueprint for circuit-based neuronal classification. *BioEssays*, *38*(10), 969–976.
<https://doi.org/10.1002/bies.201600067>
- Aso, Y., Hattori, D., Yu, Y., Johnston, R. M., Iyer, N. A., Ngo, T.-T., Dionne, H., Abbott, L., Axel, R., Tanimoto, H., & Rubin, G. M. (2014). The neuronal architecture of the mushroom body provides a logic for associative learning. *ELife*, *3*, e04577.
<https://doi.org/10.7554/eLife.04577>
- Azoulay, L., & Ramón y Cajal, S. (1894). *Les nouvelles idées sur la structure du système nerveux: Chez l'homme et chez les vertébrés* /. C. Reinwald & Cie.,
<https://doi.org/10.5962/bhl.title.48561>

- Bachman, J. A., Gyori, B. M., & Sorger, P. K. (2018). FamPlex: A resource for entity recognition and relationship resolution of human protein families and complexes in biomedical text mining. *BMC Bioinformatics*, *19*(1), 248.
<https://doi.org/10.1186/s12859-018-2211-5>
- Badea, T. C., & Nathans, J. (2011). Morphologies of mouse retinal ganglion cells expressing transcription factors Brn3a, Brn3b, and Brn3c: Analysis of wild type and mutant cells using genetically-directed sparse labeling. *Vision Research*, *51*(2), 269–279. <https://doi.org/10.1016/j.visres.2010.08.039>
- Badura, A., & De Zeeuw, C. I. (2017). Cerebellar Granule Cells: Dense, Rich and Evolving Representations. *Current Biology: CB*, *27*(11), R415–R418.
<https://doi.org/10.1016/j.cub.2017.04.009>
- Bandrowski, A. E., Cachat, J., Li, Y., Muller, H. M., Sternberg, P. W., Ciccarese, P., Clark, T., Marenco, L., Wang, R., Astakhov, V., Grethe, J. S., & Martone, M. E. (2012). A hybrid human and machine resource curation pipeline for the Neuroscience Information Framework. *Database*, *2012*(0), bas005–bas005.
<https://doi.org/10.1093/database/bas005>
- Banks, G. T., Haas, M. A., Line, S., Shepherd, H. L., AlQatari, M., Stewart, S., Rishal, I., Philpott, A., Kalmar, B., Kuta, A., Groves, M., Parkinson, N., Acevedo-Arozena, A., Brandner, S., Bannerman, D., Greensmith, L., Hafezparast, M., Koltzenburg, M., Deacon, R., ... Fisher, E. M. C. (2011). Behavioral and Other Phenotypes in a Cytoplasmic Dynein Light Intermediate Chain 1 Mutant Mouse. *Journal of*

- Neuroscience*, 31(14), 5483–5494. <https://doi.org/10.1523/JNEUROSCI.5244-10.2011>
- Barber, H. M., Ali, M. F., & Kucenas, S. (2021). Glial Patchwork: Oligodendrocyte Progenitor Cells and Astrocytes Blanket the Central Nervous System. *Frontiers in Cellular Neuroscience*, 15, 803057. <https://doi.org/10.3389/fncel.2021.803057>
- Bass, L., Clements, P., & Kazman, R. (2003). *Software Architecture in Practice (2nd Edition)*. Addison-Wesley Professional.
- Becchi, S., Buson, A., Foot, J., Jarolimek, W., & Balleine, B. W. (2017). Inhibition of semicarbazide-sensitive amine oxidase/vascular adhesion protein-1 reduces lipopolysaccharide-induced neuroinflammation. *British Journal of Pharmacology*, 174(14), 2302–2317. <https://doi.org/10.1111/bph.13832>
- Benedetti, F., Beneventano, D., Bergamaschi, S., & Simonini, G. (2019). Computing inter-document similarity with Context Semantic Analysis. *Information Systems*, 80, 136–147. <https://doi.org/10.1016/j.is.2018.02.009>
- Berzhanskaya, J., Chernyy, N., Gluckman, B. J., Schiff, S. J., & Ascoli, G. A. (2013). Modulation of hippocampal rhythms by subthreshold electric fields and network topology. *Journal of Computational Neuroscience*, 34(3), 369–389. <https://doi.org/10.1007/s10827-012-0426-4>
- Bienvenu, T. C. M., Busti, D., Magill, P. J., Ferraguti, F., & Capogna, M. (2012). Cell-Type-Specific Recruitment of Amygdala Interneurons to Hippocampal Theta Rhythm and Noxious Stimuli In Vivo. *Neuron*, 74(6), 1059–1074. <https://doi.org/10.1016/j.neuron.2012.04.022>

- Bijari, K., Akram, M. A., & Ascoli, G. A. (2020a). An open-source framework for neuroscience metadata management applied to digital reconstructions of neuronal morphology. *Brain Informatics*, 7(1), 2. <https://doi.org/10.1186/s40708-020-00103-3>
- Bijari, K., Akram, M. A., & Ascoli, G. A. (2020b). An open-source framework for neuroscience metadata management applied to digital reconstructions of neuronal morphology. *Brain Informatics*, 7(1), 2. <https://doi.org/10.1186/s40708-020-00103-3>
- Bijari, K., Akram, M. A., & Ascoli, G. A. (2020c). An open-source framework for neuroscience metadata management applied to digital reconstructions of neuronal morphology. *Brain Informatics*, 7(1), 2. <https://doi.org/10.1186/s40708-020-00103-3>
- Bijari, K., Valera, G., López-Schier, H., & Ascoli, G. A. (2021a). Quantitative neuronal morphometry by supervised and unsupervised learning. *STAR Protocols*, 2(4), 100867. <https://doi.org/10.1016/j.xpro.2021.100867>
- Bijari, K., Valera, G., López-Schier, H., & Ascoli, G. A. (2021b). Quantitative neuronal morphometry by supervised and unsupervised learning. *STAR Protocols*, 2(4), 100867. <https://doi.org/10.1016/j.xpro.2021.100867>
- Bijari, K., Zare, H., Kebriaei, E., & Veisi, H. (2020). Leveraging deep graph-based text representation for sentiment polarity applications. *Expert Systems with Applications*, 144, 113090. <https://doi.org/10.1016/j.eswa.2019.113090>

- Blanco-Suárez, E., Caldwell, A. L. M., & Allen, N. J. (2017). Role of astrocyte-synapse interactions in CNS disorders: Astrocyte-synapse disease. *The Journal of Physiology*, 595(6), 1903–1916. <https://doi.org/10.1113/JP270988>
- Blazquez-Llorca, L., Woodruff, A., Inan, M., Anderson, S. A., Yuste, R., DeFelipe, J., & Merchán-Pérez, A. (2015). Spatial distribution of neurons innervated by chandelier cells. *Brain Structure and Function*, 220(5), 2817–2834. <https://doi.org/10.1007/s00429-014-0828-3>
- Bohland, J. W., Bokil, H., Allen, C. B., & Mitra, P. P. (2009). The Brain Atlas Concordance Problem: Quantitative Comparison of Anatomical Parcellations. *PLoS ONE*, 4(9), e7200. <https://doi.org/10.1371/journal.pone.0007200>
- Bota, M., & Swanson, L. W. (2007a). The neuron classification problem. *Brain Research Reviews*, 56(1), 79–88. <https://doi.org/10.1016/j.brainresrev.2007.05.005>
- Bota, M., & Swanson, L. W. (2007b). The neuron classification problem. *Brain Research Reviews*, 56(1), 79–88. <https://doi.org/10.1016/j.brainresrev.2007.05.005>
- Bouckaert, R. (2003). Choosing between two learning algorithms based on calibrated tests. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Washington DC, 2003*. ICML, Washington DC. <https://www.cs.waikato.ac.nz/~ml/publications/2003/bouckaert-calibrated-tests.pdf>
- Bowden, D. M., Song, E., Kosheleva, J., & Dubach, M. F. (2012). NeuroNames: An Ontology for the BrainInfo Portal to Neuroscience on the Web. *Neuroinformatics*, 10(1), 97–114. <https://doi.org/10.1007/s12021-011-9128-8>

- Bragg, E. M., Fairless, E. A., Liu, S., & Briggs, F. (2017). Morphology of visual sector thalamic reticular neurons in the macaque monkey suggests retinotopically specialized, parallel stream-mixed input to the lateral geniculate nucleus. *The Journal of Comparative Neurology*, 525(5), 1273–1290.
<https://doi.org/10.1002/cne.24134>
- Breiman, L. (2001a). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Breiman, L. (2001b). Random Forests. *Machine Learning*, 45(1), 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Briggs, F., Kiley, C. W., Callaway, E. M., & Usrey, W. M. (2016). Morphological Substrates for Parallel Streams of Corticogeniculate Feedback Originating in Both V1 and V2 of the Macaque Monkey. *Neuron*, 90(2), 388–399.
<https://doi.org/10.1016/j.neuron.2016.02.038>
- Bronzuoli, M. R., Facchinetti, R., Ingrassia, D., Sarvadio, M., Schiavi, S., Steardo, L., Verkhatsky, A., Trezza, V., & Scuderi, C. (2018). Neuroglia in the autistic brain: Evidence from a preclinical model. *Molecular Autism*, 9, 66.
<https://doi.org/10.1186/s13229-018-0254-0>
- Brown, K. M., Sugihara, I., Shinoda, Y., & Ascoli, G. A. (2012). Digital Morphometry of Rat Cerebellar Climbing Fibers Reveals Distinct Branch and Bouton Types. *Journal of Neuroscience*, 32(42), 14670–14684.
<https://doi.org/10.1523/JNEUROSCI.2018-12.2012>

- Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., Larson, S. D., Rubin, D., Shepherd, G. M., Turner, J. A., & Martone, M. E. (2008). The NIFSTD and BIRNLex Vocabularies: Building Comprehensive Ontologies for Neuroscience. *Neuroinformatics*, 6(3), 175–194. <https://doi.org/10.1007/s12021-008-9032-z>
- Cannon, R. C., Turner, D. A., Pyapali, G. K., & Wheal, H. V. (1998a). An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods*, 84(1–2), 49–54. [https://doi.org/10.1016/S0165-0270\(98\)00091-0](https://doi.org/10.1016/S0165-0270(98)00091-0)
- Cannon, R. C., Turner, D. A., Pyapali, G. K., & Wheal, H. V. (1998b). An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods*, 84(1–2), 49–54. [https://doi.org/10.1016/S0165-0270\(98\)00091-0](https://doi.org/10.1016/S0165-0270(98)00091-0)
- Cannon, R. C., Turner, D. A., Pyapali, G. K., & Wheal, H. V. (1998c). An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods*, 84(1–2), 49–54. [https://doi.org/10.1016/s0165-0270\(98\)00091-0](https://doi.org/10.1016/s0165-0270(98)00091-0)
- Chiang, A.-S., Lin, C.-Y., Chuang, C.-C., Chang, H.-M., Hsieh, C.-H., Yeh, C.-W., Shih, C.-T., Wu, J.-J., Wang, G.-T., Chen, Y.-C., Wu, C.-C., Chen, G.-Y., Ching, Y.-T., Lee, P.-C., Lin, C.-Y., Lin, H.-H., Wu, C.-C., Hsu, H.-W., Huang, Y.-A., ... Hwang, J.-K. (2011). Three-Dimensional Reconstruction of Brain-wide Wiring Networks in *Drosophila* at Single-Cell Resolution. *Current Biology*, 21(1), 1–11. <https://doi.org/10.1016/j.cub.2010.11.056>
- Chu, P., Peck, J., & Brumberg, J. C. (2015a). Exercises in Anatomy, Connectivity, and Morphology using Neuromorpho.org and the Allen Brain Atlas. *Journal of*

Undergraduate Neuroscience Education: JUNE: A Publication of FUN, Faculty for Undergraduate Neuroscience, 13(2), A95–A100.

Chu, P., Peck, J., & Brumberg, J. C. (2015b). Exercises in Anatomy, Connectivity, and Morphology using Neuromorpho.org and the Allen Brain Atlas. *Journal of Undergraduate Neuroscience Education: JUNE: A Publication of FUN, Faculty for Undergraduate Neuroscience, 13(2), A95–A100.*

Conjeti, S., Mesbah, S., Negahdar, M., Rautenberg, P. L., Zhang, S., Navab, N., & Katouzian, A. (2016a). Neuron-Miner: An Advanced Tool for Morphological Search and Retrieval in Neuroscientific Image Databases. *Neuroinformatics, 14(4), 369–385.* <https://doi.org/10.1007/s12021-016-9300-2>

Conjeti, S., Mesbah, S., Negahdar, M., Rautenberg, P. L., Zhang, S., Navab, N., & Katouzian, A. (2016b). Neuron-Miner: An Advanced Tool for Morphological Search and Retrieval in Neuroscientific Image Databases. *Neuroinformatics, 14(4), 369–385.* <https://doi.org/10.1007/s12021-016-9300-2>

Core Team, R. (2021). *R: A Language and Environment for Statistical Computing* [R]. R Foundation for Statistical Computing. <https://www.R-project.org/>

Cortes, C., & Vapnik, V. (1995a). Support-vector networks. *Machine Learning, 20(3), 273–297.* <https://doi.org/10.1007/BF00994018>

Cortes, C., & Vapnik, V. (1995b). Support-vector networks. *Machine Learning, 20(3), 273–297.* <https://doi.org/10.1007/BF00994018>

Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S., & Jefferis, G. S. X. E. (2016a). NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction

- of Neuron Family Databases. *Neuron*, 91(2), 293–311.
<https://doi.org/10.1016/j.neuron.2016.06.012>
- Costa, M., Manton, J. D., Ostrovsky, A. D., Prohaska, S., & Jefferis, G. S. X. E. (2016b). NBLAST: Rapid, Sensitive Comparison of Neuronal Structure and Construction of Neuron Family Databases. *Neuron*, 91(2), 293–311.
<https://doi.org/10.1016/j.neuron.2016.06.012>
- Cuntz, H. (2012). The Dendritic Density Field of a Cortical Pyramidal Cell. *Frontiers in Neuroanatomy*, 6. <https://doi.org/10.3389/fnana.2012.00002>
- Cuntz, H., Forstner, F., Borst, A., & Häusser, M. (2010). One Rule to Grow Them All: A General Theory of Neuronal Branching and Its Practical Application. *PLoS Computational Biology*, 6(8), e1000877.
<https://doi.org/10.1371/journal.pcbi.1000877>
- Cuntz, H., Forstner, F., Borst, A., & Häusser, M. (2011). The TREES Toolbox—Probing the Basis of Axonal and Dendritic Branching. *Neuroinformatics*, 9(1), 91–96.
<https://doi.org/10.1007/s12021-010-9093-7>
- Dayan, P. (1994). Computational modelling. *Current Opinion in Neurobiology*, 4(2), 212–217. [https://doi.org/10.1016/0959-4388\(94\)90075-2](https://doi.org/10.1016/0959-4388(94)90075-2)
- Defelipe, J., Hendry, S. H. C., Jones, E. G., & Schmechel, D. (1985). Variability in the terminations of GABAergic chandelier cell axons on initial segments of pyramidal cell axons in the monkey sensory-motor cortex. *The Journal of Comparative Neurology*, 231(3), 364–384. <https://doi.org/10.1002/cne.902310307>

- Di Bucchianico, A. (2008). Coefficient of Determination (R^2). In F. Ruggeri, R. S. Kenett, & F. W. Faltin (Eds.), *Encyclopedia of Statistics in Quality and Reliability* (p. eqr173). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9780470061572.eqr173>
- Donnelly, C. R., Chen, O., & Ji, R.-R. (2020). How Do Sensory Neurons Sense Danger Signals? *Trends in Neurosciences*, *43*(10), 822–838.
<https://doi.org/10.1016/j.tins.2020.07.008>
- Duan, H. (2003). Age-related Dendritic and Spine Changes in Corticocortically Projecting Neurons in Macaque Monkeys. *Cerebral Cortex*, *13*(9), 950–961.
<https://doi.org/10.1093/cercor/13.9.950>
- Duan, H., Wearne, S. L., Morrison, J. H., & Hof, P. R. (2002a). Quantitative analysis of the dendritic morphology of corticocortical projection neurons in the macaque monkey association cortex. *Neuroscience*, *114*(2), 349–359.
[https://doi.org/10.1016/S0306-4522\(02\)00305-6](https://doi.org/10.1016/S0306-4522(02)00305-6)
- Duan, H., Wearne, S. L., Morrison, J. H., & Hof, P. R. (2002b). Quantitative analysis of the dendritic morphology of corticocortical projection neurons in the macaque monkey association cortex. *Neuroscience*, *114*(2), 349–359.
[https://doi.org/10.1016/S0306-4522\(02\)00305-6](https://doi.org/10.1016/S0306-4522(02)00305-6)
- Dulamea, A. O. (2017). Role of Oligodendrocyte Dysfunction in Demyelination, Remyelination and Neurodegeneration in Multiple Sclerosis. In A. A. A. Asea, F. Geraci, & P. Kaur (Eds.), *Multiple Sclerosis: Bench to Bedside* (Vol. 958, pp. 91–

- 127). Springer International Publishing. https://doi.org/10.1007/978-3-319-47861-6_7
- Dzyubenko, E., Gottschling, C., & Faissner, A. (2016). Neuron-Glia Interactions in Neural Plasticity: Contributions of Neural Extracellular Matrix and Perineuronal Nets. *Neural Plasticity*, 2016, 1–14. <https://doi.org/10.1155/2016/5214961>
- Economo, M. N., Clack, N. G., Lavis, L. D., Gerfen, C. R., Svoboda, K., Myers, E. W., & Chandrashekar, J. (2016). A platform for brain-wide imaging and reconstruction of individual neurons. *ELife*, 5, e10566. <https://doi.org/10.7554/eLife.10566>
- Edwards, J., Daniel, E., Kinney, J., Bartol, T., Sejnowski, T., Johnston, D., Harris, K., & Bajaj, C. (2014). VolRoverN: Enhancing surface and volumetric reconstruction for realistic dynamical simulation of cellular and subcellular function. *Neuroinformatics*, 12(2), 277–289. <https://doi.org/10.1007/s12021-013-9205-2>
- Egyedi, A. L., O’Connor, M. J., Martínez-Romero, M., Willrett, D., Hardi, J., Graybeal, J., & Musen, M. A. (2018). *Using Semantic Technologies to Enhance Metadata Submissions to Public Repositories in Biomedicine*. 930325 Bytes. <https://doi.org/10.6084/M9.FIGSHARE.7324175>
- Euler, T., Haverkamp, S., Schubert, T., & Baden, T. (2014). Retinal bipolar cells: Elementary building blocks of vision. *Nature Reviews Neuroscience*, 15(8), 507–519. <https://doi.org/10.1038/nrn3783>
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and

- weaknesses. *The FASEB Journal*, 22(2), 338–342. <https://doi.org/10.1096/fj.07-9492LSF>
- Favero, M., & Castro-Alamancos, M. A. (2013). Synaptic Cooperativity Regulates Persistent Network Activity in Neocortex. *Journal of Neuroscience*, 33(7), 3151–3163. <https://doi.org/10.1523/JNEUROSCI.4424-12.2013>
- Feng, L., Zhao, T., & Kim, J. (2015). neuTube 1.0: A New Design for Efficient Neuron Reconstruction Software Based on the SWC Format. *Eneuro*, 2(1), ENEURO.0049-14.2014. <https://doi.org/10.1523/ENEURO.0049-14.2014>
- Fenno, L. E., Mattis, J., Ramakrishnan, C., Hyun, M., Lee, S. Y., He, M., Tucciarone, J., Selimbeyoglu, A., Berndt, A., Grosenick, L., Zalocusky, K. A., Bernstein, H., Swanson, H., Perry, C., Diester, I., Boyce, F. M., Bass, C. E., Neve, R., Huang, Z. J., & Deisseroth, K. (2014). Targeting cells with single vectors using multiple-feature Boolean logic. *Nature Methods*, 11(7), 763–772. <https://doi.org/10.1038/nmeth.2996>
- Friedman, R. (2020). Tools to automate use of the database at NeuroMorpho.Org. *Software Impacts*, 6, 100032. <https://doi.org/10.1016/j.simpa.2020.100032>
- Fujishima, K., Kawabata Galbraith, K., & Kengaku, M. (2018). Dendritic Self-Avoidance and Morphological Development of Cerebellar Purkinje Cells. *Cerebellum (London, England)*, 17(6), 701–708. <https://doi.org/10.1007/s12311-018-0984-8>
- Gang, Y., Zhou, H., Jia, Y., Liu, L., Liu, X., Rao, G., Li, L., Wang, X., Lv, X., Xiong, H., Yang, Z., Luo, Q., Gong, H., & Zeng, S. (2017). Embedding and Chemical Reactivation of Green Fluorescent Protein in the Whole Mouse Brain for Optical

- Micro-Imaging. *Frontiers in Neuroscience*, 11.
<https://doi.org/10.3389/fnins.2017.00121>
- Ganglberger, F., Schulze, F., Tirian, L., Novikov, A., Dickson, B., Bühler, K., & Langs, G. (2014). Structure-Based Neuron Retrieval Across *Drosophila* Brains. *Neuroinformatics*, 12(3), 423–434. <https://doi.org/10.1007/s12021-014-9219-4>
- García-Marín, V., García-López, P., & Freire, M. (2007a). Cajal's contributions to glia research. *Trends in Neurosciences*, 30(9), 479–487.
<https://doi.org/10.1016/j.tins.2007.06.008>
- García-Marín, V., García-López, P., & Freire, M. (2007b). Cajal's contributions to glia research. *Trends in Neurosciences*, 30(9), 479–487.
<https://doi.org/10.1016/j.tins.2007.06.008>
- Gardner, D., Goldberg, D. H., Grafstein, B., Robert, A., & Gardner, E. P. (2008). Terminology for Neuroscience Data Discovery: Multi-tree Syntax and Investigator-Derived Semantics. *Neuroinformatics*, 6(3), 161–174.
<https://doi.org/10.1007/s12021-008-9029-7>
- Ghosh, S. K. (2020). Camillo Golgi (1843-1926): Scientist extraordinaire and pioneer figure of modern neurology. *Anatomy & Cell Biology*, 53(4), 385–392.
<https://doi.org/10.5115/acb.20.196>
- Gillette, T. A., Brown, K. M., & Ascoli, G. A. (2011). The DIADEM metric: Comparing multiple reconstructions of the same neuron. *Neuroinformatics*, 9(2–3), 233–245.
<https://doi.org/10.1007/s12021-011-9117-y>

- Gillette, T. A., Hosseini, P., & Ascoli, G. A. (2015). Topological characterization of neuronal arbor morphology via sequence representation: II - global alignment. *BMC Bioinformatics*, *16*(1), 209. <https://doi.org/10.1186/s12859-015-0605-1>
- Gleeson, P., Cantarelli, M., Marin, B., Quintana, A., Earnshaw, M., Sadeh, S., Piasini, E., Birgiolas, J., Cannon, R. C., Cayco-Gajic, N. A., Crook, S., Davison, A. P., Dura-Bernal, S., Ecker, A., Hines, M. L., Idili, G., Lanore, F., Larson, S. D., Lytton, W. W., ... Silver, R. A. (2019). Open Source Brain: A Collaborative Resource for Visualizing, Analyzing, Simulating, and Developing Standardized Models of Neurons and Circuits. *Neuron*, *103*(3), 395-411.e5. <https://doi.org/10.1016/j.neuron.2019.05.019>
- Gleeson, P., Crook, S., Cannon, R. C., Hines, M. L., Billings, G. O., Farinella, M., Morse, T. M., Davison, A. P., Ray, S., Bhalla, U. S., Barnes, S. R., Dimitrova, Y. D., & Silver, R. A. (2010). NeuroML: A Language for Describing Data Driven Models of Neurons and Networks with a High Degree of Biological Detail. *PLoS Computational Biology*, *6*(6), e1000815. <https://doi.org/10.1371/journal.pcbi.1000815>
- Gleeson, P., Davison, A. P., Silver, R. A., & Ascoli, G. A. (2017a). A Commitment to Open Source in Neuroscience. *Neuron*, *96*(5), 964–965. <https://doi.org/10.1016/j.neuron.2017.10.013>
- Gleeson, P., Davison, A. P., Silver, R. A., & Ascoli, G. A. (2017b). A Commitment to Open Source in Neuroscience. *Neuron*, *96*(5), 964–965. <https://doi.org/10.1016/j.neuron.2017.10.013>

- Goddard, N. H., Hucka, M., Howell, F., Cornelis, H., Shankar, K., & Beeman, D. (2001). Towards NeuroML: Model Description Methods for Collaborative Modelling in Neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1412), 1209–1228.
<https://doi.org/10.1098/rstb.2001.0910>
- Gonçalves, R. S., O'Connor, M. J., Martínez-Romero, M., Egyedi, A. L., Willrett, D., Graybeal, J., & Musen, M. A. (2017). The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments. In C. d'Amato, M. Fernandez, V. Tamma, F. Lecue, P. Cudré-Mauroux, J. Sequeda, C. Lange, & J. Heflin (Eds.), *The Semantic Web – ISWC 2017* (Vol. 10588, pp. 103–110). Springer International Publishing.
https://doi.org/10.1007/978-3-319-68204-4_10
- Gong, H., Xu, D., Yuan, J., Li, X., Guo, C., Peng, J., Li, Y., Schwarz, L. A., Li, A., Hu, B., Xiong, B., Sun, Q., Zhang, Y., Liu, J., Zhong, Q., Xu, T., Zeng, S., & Luo, Q. (2016). High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level. *Nature Communications*, 7(1), 12142. <https://doi.org/10.1038/ncomms12142>
- Gong, H., Zeng, S., Yan, C., Lv, X., Yang, Z., Xu, T., Feng, Z., Ding, W., Qi, X., Li, A., Wu, J., & Luo, Q. (2013). Continuously tracing brain-wide long-distance axonal projections in mice at a one-micron voxel resolution. *NeuroImage*, 74, 87–98.
<https://doi.org/10.1016/j.neuroimage.2013.02.005>

- Greggio, E., Bubacco, L., & Russo, I. (2017). Cross-talk between LRRK2 and PKA: Implication for Parkinson's disease? *Biochemical Society Transactions*, *45*(1), 261–267. <https://doi.org/10.1042/BST20160396>
- Grewe, J., Wachtler, T., & Benda, J. (2011). A Bottom-up Approach to Data Annotation in Neurophysiology. *Frontiers in Neuroinformatics*, *5*. <https://doi.org/10.3389/fninf.2011.00016>
- Halavi, M., Hamilton, K. A., Parekh, R., & Ascoli, G. A. (2012a). Digital Reconstructions of Neuronal Morphology: Three Decades of Research Trends. *Frontiers in Neuroscience*, *6*. <https://doi.org/10.3389/fnins.2012.00049>
- Halavi, M., Hamilton, K. A., Parekh, R., & Ascoli, G. A. (2012b). Digital Reconstructions of Neuronal Morphology: Three Decades of Research Trends. *Frontiers in Neuroscience*, *6*. <https://doi.org/10.3389/fnins.2012.00049>
- Halavi, M., Hamilton, K. A., Parekh, R., & Ascoli, G. A. (2012c). Digital Reconstructions of Neuronal Morphology: Three Decades of Research Trends. *Frontiers in Neuroscience*, *6*. <https://doi.org/10.3389/fnins.2012.00049>
- Halavi, M., Polavaram, S., Donohue, D. E., Hamilton, G., Hoyt, J., Smith, K. P., & Ascoli, G. A. (2008a). NeuroMorpho.Org Implementation of Digital Neuroscience: Dense Coverage and Integration with the NIF. *Neuroinformatics*, *6*(3), 241. <https://doi.org/10.1007/s12021-008-9030-1>
- Halavi, M., Polavaram, S., Donohue, D. E., Hamilton, G., Hoyt, J., Smith, K. P., & Ascoli, G. A. (2008b). NeuroMorpho.Org implementation of digital neuroscience:

- Dense coverage and integration with the NIF. *Neuroinformatics*, 6(3), 241–252.
<https://doi.org/10.1007/s12021-008-9030-1>
- Hamilton, D. J., Shepherd, G. M., Martone, M. E., & Ascoli, G. A. (2012). An ontological approach to describing neurons and their relationships. *Frontiers in Neuroinformatics*, 6. <https://doi.org/10.3389/fninf.2012.00015>
- Hamilton, D. J., Wheeler, D. W., White, C. M., Rees, C. L., Komendantov, A. O., Bergamino, M., & Ascoli, G. A. (2017). Name-calling in the hippocampus (and beyond): Coming to terms with neuron types and properties. *Brain Informatics*, 4(1), 1–12. <https://doi.org/10.1007/s40708-016-0053-3>
- Harrel, F., & Dupont, C. (2021). *Harrell Miscellaneous* (4.6-0) [R].
<https://hbiostat.org/R/Hmisc/>
- Harris, J. A., Hirokawa, K. E., Sorensen, S. A., Gu, H., Mills, M., Ng, L. L., Bohn, P., Mortrud, M., Ouellette, B., Kidney, J., Smith, K. A., Dang, C., Sunkin, S., Bernard, A., Oh, S. W., Madisen, L., & Zeng, H. (2014). Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Frontiers in Neural Circuits*, 8. <https://doi.org/10.3389/fncir.2014.00076>
- Harris, K. D., & Shepherd, G. M. G. (2015). The neocortical circuit: Themes and variations. *Nature Neuroscience*, 18(2), 170–181. <https://doi.org/10.1038/nn.3917>
- He, M., Tucciarone, J., Lee, S., Nigro, M. J., Kim, Y., Levine, J. M., Kelly, S. M., Krugikov, I., Wu, P., Chen, Y., Gong, L., Hou, Y., Osten, P., Rudy, B., & Huang, Z. J. (2016). Strategies and Tools for Combinatorial Targeting of GABAergic

- Neurons in Mouse Cerebral Cortex. *Neuron*, 91(6), 1228–1243.
<https://doi.org/10.1016/j.neuron.2016.08.021>
- Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013a). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), 168–174. <https://doi.org/10.1038/nature12346>
- Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013b). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), 168–174. <https://doi.org/10.1038/nature12346>
- Herculano-Houzel, S. (2010). Coordinated scaling of cortical and cerebellar numbers of neurons. *Frontiers in Neuroanatomy*, 4, 12.
<https://doi.org/10.3389/fnana.2010.00012>
- Hines, M. L., & Carnevale, N. T. (2001). Neuron: A Tool for Neuroscientists. *The Neuroscientist*, 7(2), 123–135. <https://doi.org/10.1177/107385840100700207>
- Hirano, T. (2018). Purkinje Neurons: Development, Morphology, and Function. *Cerebellum (London, England)*, 17(6), 699–700. <https://doi.org/10.1007/s12311-018-0985-7>
- Hobert, O., Glenwinkel, L., & White, J. (2016). Revisiting Neuronal Cell Type Classification in *Caenorhabditis elegans*. *Current Biology*, 26(22), R1197–R1203.
<https://doi.org/10.1016/j.cub.2016.10.027>
- Holt Zaugg, Richard E. West, Isaku Tateishi, Daniel L. Randall. (2011). Mendeley: Creating Communities of Scholarly Inquiry Through Research Collaboration. *TechTrends*, 55(1), 32–36. <https://doi.org/10.1007/s11528-011-0467-y>

- Houston, C. M., Diamanti, E., Diamantaki, M., Kutsarova, E., Cook, A., Sultan, F., & Brickley, S. G. (2017). Exploring the significance of morphological diversity for cerebellar granule cell excitability. *Scientific Reports*, *7*, 46147. <https://doi.org/10.1038/srep46147>
- Huang, Z. J., & Zeng, H. (2013). Genetic approaches to neural circuits in the mouse. *Annual Review of Neuroscience*, *36*, 183–215. <https://doi.org/10.1146/annurev-neuro-062012-170307>
- Hutchins, B. I., Baker, K. L., Davis, M. T., Diwersy, M. A., Haque, E., Harriman, R. M., Hoppe, T. A., Leicht, S. A., Meyer, P., & Santangelo, G. M. (2019). The NIH Open Citation Collection: A public access, broad coverage resource. *PLOS Biology*, *17*(10), e3000385. <https://doi.org/10.1371/journal.pbio.3000385>
- Insel, T. R., Landis, S. C., & Collins, F. S. (2013). The NIH BRAIN Initiative. *Science*, *340*(6133), 687–688. <https://doi.org/10.1126/science.1239276>
- Ishizuka, N., Cowan, W. M., & Amaral, D. G. (1995). A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *The Journal of Comparative Neurology*, *362*(1), 17–45. <https://doi.org/10.1002/cne.903620103>
- J. Rodríguez, J., M. Butt, A., Gardenal, E., Parpura, V., & Verkhratsky, A. (2016). Complex and Differential Glial Responses in Alzheimers Disease and Ageing. *Current Alzheimer Research*, *13*(4), 343–358. <https://doi.org/10.2174/1567205013666160229112911>

- Jackson, D. A. (1993). Stopping Rules in Principal Components Analysis: A Comparison of Heuristical and Statistical Approaches. *Ecology*, 74(8), 2204–2214.
<https://doi.org/10.2307/1939574>
- Jacobs, G. A., & Theunissen, F. E. (2000). Extraction of Sensory Parameters from a Neural Map by Primary Sensory Interneurons. *The Journal of Neuroscience*, 20(8), 2934–2943. <https://doi.org/10.1523/JNEUROSCI.20-08-02934.2000>
- Jefferis, G. S. X. E., Potter, C. J., Chan, A. M., Marin, E. C., Rohlfsing, T., Maurer, C. R., & Luo, L. (2007). Comprehensive Maps of Drosophila Higher Olfactory Centers: Spatially Segregated Fruit and Pheromone Representation. *Cell*, 128(6), 1187–1203. <https://doi.org/10.1016/j.cell.2007.01.040>
- Jenett, A., Rubin, G. M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, J. H., Peng, H., Trautman, E. T., Svirskas, R. R., Myers, E. W., Iwinski, Z. R., Aso, Y., ... Zugates, C. T. (2012). A GAL4-Driver Line Resource for Drosophila Neurobiology. *Cell Reports*, 2(4), 991–1001.
<https://doi.org/10.1016/j.celrep.2012.09.011>
- Jessen, K. R. (2004). Glial cells. *The International Journal of Biochemistry & Cell Biology*, 36(10), 1861–1867. <https://doi.org/10.1016/j.biocel.2004.02.023>
- Johnson, D., & Sinanovic, S. (2001). *Symmetrizing the Kullback-Leibler Distance*.
<https://hdl.handle.net/1911/19969>

- Johnson, J., Douze, M., & Jegou, H. (2021). Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535–547.
<https://doi.org/10.1109/TBDATA.2019.2921572>
- Kanari, L., Dłotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., & Markram, H. (2018a). A Topological Representation of Branching Neuronal Morphologies. *Neuroinformatics*, 16(1), 3–13. <https://doi.org/10.1007/s12021-017-9341-1>
- Kanari, L., Dłotko, P., Scolamiero, M., Levi, R., Shillcock, J., Hess, K., & Markram, H. (2018b). A Topological Representation of Branching Neuronal Morphologies. *Neuroinformatics*, 16(1), 3–13. <https://doi.org/10.1007/s12021-017-9341-1>
- Karelina, K., Sarac, B., Freeman, L. M., Gaier, K. R., & Weil, Z. M. (2016). Traumatic brain injury and obesity induce persistent central insulin resistance. *The European Journal of Neuroscience*, 43(8), 1034–1043. <https://doi.org/10.1111/ejn.13194>
- Kawaguchi, Y., Karube, F., & Kubota, Y. (2006). Dendritic branch typing and spine expression patterns in cortical nonpyramidal cells. *Cerebral Cortex (New York, N.Y.: 1991)*, 16(5), 696–711. <https://doi.org/10.1093/cercor/bhj015>
- Kawaguchi, Y., & Shindou, T. (1998). Noradrenergic Excitation and Inhibition of GABAergic Cell Types in Rat Frontal Cortex. *The Journal of Neuroscience*, 18(17), 6963–6976. <https://doi.org/10.1523/JNEUROSCI.18-17-06963.1998>
- Khakh, B. S., & Deneen, B. (2019). The Emerging Nature of Astrocyte Diversity. *Annual Review of Neuroscience*, 42, 187–207. <https://doi.org/10.1146/annurev-neuro-070918-050443>

- Koene, R. A., Tijms, B., van Hees, P., Postma, F., de Ridder, A., Ramakers, G. J. A., van Pelt, J., & van Ooyen, A. (2009). NETMORPH: A framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. *Neuroinformatics*, 7(3), 195–210. <https://doi.org/10.1007/s12021-009-9052-3>
- Kohli, P. S., & Arora, S. (2018). Application of Machine Learning in Disease Prediction. *2018 4th International Conference on Computing Communication and Automation (ICCCA)*, 1–4. <https://doi.org/10.1109/CCAA.2018.8777449>
- Koopmans, F., van Nierop, P., Andres-Alonso, M., Byrnes, A., Cijssouw, T., Coba, M. P., Cornelisse, L. N., Farrell, R. J., Goldschmidt, H. L., Howrigan, D. P., Hussain, N. K., Imig, C., de Jong, A. P. H., Jung, H., Kohansalnodehi, M., Kramarz, B., Lipstein, N., Lovering, R. C., MacGillavry, H., ... Verhage, M. (2019). SynGO: An Evidence-Based, Expert-Curated Knowledge Base for the Synapse. *Neuron*, 103(2), 217-234.e4. <https://doi.org/10.1016/j.neuron.2019.05.002>
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26(3), 159–190. <https://doi.org/10.1007/s10462-007-9052-3>
- Krashia, P., Cordella, A., Nobili, A., La Barbera, L., Federici, M., Leuti, A., Campanelli, F., Natale, G., Marino, G., Calabrese, V., Vedele, F., Ghiglieri, V., Picconi, B., Di Lazzaro, G., Schirinzi, T., Sancesario, G., Casadei, N., Riess, O., Bernardini, S., ... Mercuri, N. B. (2019). Blunting neuroinflammation with resolvin D1 prevents early pathology in a rat model of Parkinson's disease. *Nature Communications*, 10(1), 3945. <https://doi.org/10.1038/s41467-019-11928-w>

- Kubota, Y. (2014). Untangling GABAergic wiring in the cortical microcircuit. *Current Opinion in Neurobiology*, 26, 7–14. <https://doi.org/10.1016/j.conb.2013.10.003>
- Kuhn, M. (2021). *caret: Classification and Regression Training (6.0-90)* [R].
<https://CRAN.R-project.org/package=caret>
- Larson, S. D., & Martone, M. E. (2013). NeuroLex.org: An online framework for neuroscience knowledge. *Frontiers in Neuroinformatics*, 7.
<https://doi.org/10.3389/fninf.2013.00018>
- Laternus, S., Kobak, D., & Berens, P. (2020). A Systematic Evaluation of Interneuron Morphology Representations for Cell Type Discrimination. *Neuroinformatics*, 18(4), 591–609. <https://doi.org/10.1007/s12021-020-09461-z>
- Lazarewicz, M. T., Migliore, M., & Ascoli, G. A. (2002). A new bursting model of CA3 pyramidal cell physiology suggests multiple locations for spike initiation. *Bio Systems*, 67(1–3), 129–137. [https://doi.org/10.1016/s0303-2647\(02\)00071-0](https://doi.org/10.1016/s0303-2647(02)00071-0)
- Ledderose, J., Senci3n, L., Salgado, H., Arias-Carri3n, O., & Trevi3no, M. (2014). A software tool for the analysis of neuronal morphology data. *International Archives of Medicine*, 7(1), 6. <https://doi.org/10.1186/1755-7682-7-6>
- Lewis, D. A., & Lund, J. S. (1990). Heterogeneity of chandelier neurons in monkey neocortex: Corticotropin-releasing factor- and parvalbumin-immunoreactive populations. *The Journal of Comparative Neurology*, 293(4), 599–615.
<https://doi.org/10.1002/cne.902930406>
- Leyh, J., Paeschke, S., Mages, B., Michalski, D., Nowicki, M., Bechmann, I., & Winter, K. (2021). Classification of Microglial Morphological Phenotypes Using Machine

- Learning. *Frontiers in Cellular Neuroscience*, *15*, 701673.
<https://doi.org/10.3389/fncel.2021.701673>
- Li, A., Gong, H., Zhang, B., Wang, Q., Yan, C., Wu, J., Liu, Q., Zeng, S., & Luo, Q. (2010). Micro-Optical Sectioning Tomography to Obtain a High-Resolution Atlas of the Mouse Brain. *Science*, *330*(6009), 1404–1408.
<https://doi.org/10.1126/science.1191776>
- Li, Y., Brewer, D., Burke, R. E., & Ascoli, G. A. (2005). Developmental changes in spinal motoneuron dendrites in neonatal mice. *The Journal of Comparative Neurology*, *483*(3), 304–317. <https://doi.org/10.1002/cne.20438>
- Li, Y., Gong, H., Yang, X., Yuan, J., Jiang, T., Li, X., Sun, Q., Zhu, D., Wang, Z., Luo, Q., & Li, A. (2017). TDat: An Efficient Platform for Processing Petabyte-Scale Whole-Brain Volumetric Images. *Frontiers in Neural Circuits*, *11*, 51.
<https://doi.org/10.3389/fncir.2017.00051>
- Li, Y., Wang, D., Ascoli, G. A., Mitra, P., & Wang, Y. (2017a). Metrics for comparing neuronal tree shapes based on persistent homology. *PLOS ONE*, *12*(8), e0182184.
<https://doi.org/10.1371/journal.pone.0182184>
- Li, Y., Wang, D., Ascoli, G. A., Mitra, P., & Wang, Y. (2017b). Metrics for comparing neuronal tree shapes based on persistent homology. *PLOS ONE*, *12*(8), e0182184.
<https://doi.org/10.1371/journal.pone.0182184>
- Li, Z., Fang, R., Shen, F., Katouzian, A., & Zhang, S. (2017). Indexing and mining large-scale neuron databases using maximum inner product search. *Pattern Recognition*, *63*, 680–688. <https://doi.org/10.1016/j.patcog.2016.09.041>

- Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest* [R].
<https://CRAN.R-project.org/doc/Rnews/>
- Lichtman, J. W., & Denk, W. (2011). The Big and the Small: Challenges of Imaging the Brain's Circuits. *Science*, *334*(6056), 618–623.
<https://doi.org/10.1126/science.1209168>
- Liddel, S. A., Guttenplan, K. A., Clarke, L. E., Bennett, F. C., Bohlen, C. J., Schirmer, L., Bennett, M. L., Münch, A. E., Chung, W.-S., Peterson, T. C., Wilton, D. K., Frouin, A., Napier, B. A., Panicker, N., Kumar, M., Buckwalter, M. S., Rowitch, D. H., Dawson, V. L., Dawson, T. M., ... Barres, B. A. (2017). Neurotoxic reactive astrocytes are induced by activated microglia. *Nature*, *541*(7638), 481–487. <https://doi.org/10.1038/nature21029>
- Litvina, E., Adams, A., Barth, A., Bruchez, M., Carson, J., Chung, J. E., Dupre, K. B., Frank, L. M., Gates, K. M., Harris, K. M., Joo, H., William Lichtman, J., Ramos, K. M., Sejnowski, T., Trimmer, J. S., White, S., & Koroshetz, W. (2019a). BRAIN Initiative: Cutting-Edge Tools and Resources for the Community. *The Journal of Neuroscience*, *39*(42), 8275–8284.
<https://doi.org/10.1523/JNEUROSCI.1169-19.2019>
- Litvina, E., Adams, A., Barth, A., Bruchez, M., Carson, J., Chung, J. E., Dupre, K. B., Frank, L. M., Gates, K. M., Harris, K. M., Joo, H., William Lichtman, J., Ramos, K. M., Sejnowski, T., Trimmer, J. S., White, S., & Koroshetz, W. (2019b). BRAIN Initiative: Cutting-Edge Tools and Resources for the Community. *The*

- Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 39(42), 8275–8284. <https://doi.org/10.1523/JNEUROSCI.1169-19.2019>
- Ljungquist, B., Akram, M. A., & Ascoli, G. A. (2021). *Large scale similarity search across digital reconstructions of neural morphology* [Preprint]. *Neuroscience*. <https://doi.org/10.1101/2021.12.17.473026>
- Longo, D. L., & Drazen, J. M. (2016). Data Sharing. *New England Journal of Medicine*, 374(3), 276–277. <https://doi.org/10.1056/NEJMe1516564>
- López-Cruz, P. L., Bielza, C., Larrañaga, P., Benavides-Piccione, R., & DeFelipe, J. (2011a). Models and Simulation of 3D Neuronal Dendritic Trees Using Bayesian Networks. *Neuroinformatics*, 9(4), 347–369. <https://doi.org/10.1007/s12021-011-9103-4>
- López-Cruz, P. L., Bielza, C., Larrañaga, P., Benavides-Piccione, R., & DeFelipe, J. (2011b). Models and Simulation of 3D Neuronal Dendritic Trees Using Bayesian Networks. *Neuroinformatics*, 9(4), 347–369. <https://doi.org/10.1007/s12021-011-9103-4>
- Lu, J. (2011). Neuronal tracing for connectomic studies. *Neuroinformatics*, 9(2–3), 159–166. <https://doi.org/10.1007/s12021-011-9101-6>
- Lu, J., Tucciarone, J., Padilla-Coreano, N., He, M., Gordon, J. A., & Huang, Z. J. (2017). Selective inhibitory control of pyramidal neuron ensembles and cortical subnetworks by chandelier cells. *Nature Neuroscience*, 20(10), 1377–1383. <https://doi.org/10.1038/nn.4624>

- Lu, Y., Carin, L., Coifman, R., Shain, W., & Roysam, B. (2015a). Quantitative Arbor Analytics: Unsupervised Harmonic Co-Clustering of Populations of Brain Cell Arbors Based on L-Measure. *Neuroinformatics*, *13*(1), 47–63.
<https://doi.org/10.1007/s12021-014-9237-2>
- Lu, Y., Carin, L., Coifman, R., Shain, W., & Roysam, B. (2015b). Quantitative arbor analytics: Unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. *Neuroinformatics*, *13*(1), 47–63.
<https://doi.org/10.1007/s12021-014-9237-2>
- Lu, Y., Carin, L., Coifman, R., Shain, W., & Roysam, B. (2015c). Quantitative arbor analytics: Unsupervised harmonic co-clustering of populations of brain cell arbors based on L-measure. *Neuroinformatics*, *13*(1), 47–63.
<https://doi.org/10.1007/s12021-014-9237-2>
- Luts, J., Ojeda, F., Van de Plas, R., De Moor, B., Van Huffel, S., & Suykens, J. A. K. (2010). A tutorial on support vector machine-based methods for classification problems in chemometrics. *Analytica Chimica Acta*, *665*(2), 129–145.
<https://doi.org/10.1016/j.aca.2010.03.030>
- Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., Trombetta, J. J., Weitz, D. A., Sanes, J. R., Shalek, A. K., Regev, A., & McCarroll, S. A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell*, *161*(5), 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>

- Magalhães, B. R. C., Sterling, T., Hines, M., & Schürmann, F. (2019). Asynchronous Branch-Parallel Simulation of Detailed Neuron Models. *Frontiers in Neuroinformatics, 13*, 54. <https://doi.org/10.3389/fninf.2019.00054>
- Maraver, P., Armañanzas, R., Gillette, T. A., & Ascoli, G. A. (2019a). PaperBot: Open-source web-based search and metadata organization of scientific literature. *BMC Bioinformatics, 20*(1), 50. <https://doi.org/10.1186/s12859-019-2613-z>
- Maraver, P., Armañanzas, R., Gillette, T. A., & Ascoli, G. A. (2019b). PaperBot: Open-source web-based search and metadata organization of scientific literature. *BMC Bioinformatics, 20*(1), 50. <https://doi.org/10.1186/s12859-019-2613-z>
- Markram, H. (2012). The human brain project. *Scientific American, 306*(6), 50–55. <https://doi.org/10.1038/scientificamerican0612-50>
- Martinez, L. M., Wang, Q., Reid, R. C., Pillai, C., Alonso, J.-M., Sommer, F. T., & Hirsch, J. A. (2005). Receptive field structure varies with layer in the primary visual cortex. *Nature Neuroscience, 8*(3), 372–379. <https://doi.org/10.1038/nn1404>
- Martínez-Romero, M., O'Connor, M. J., Egyedi, A. L., Willrett, D., Hardi, J., Graybeal, J., & Musen, M. A. (2019). Using association rule mining and ontologies to generate metadata recommendations from multiple biomedical databases. *Database, 2019*, baz059. <https://doi.org/10.1093/database/baz059>
- Matas, S. L. de A., Glehn, F. von, Fernandes, G. B. P., & Soares, C. A. S. (2013). Cerebrospinal fluid analysis in the context of CNS demyelinating diseases.

- Arquivos De Neuro-Psiquiatria*, 71(9B), 685–688. <https://doi.org/10.1590/0004-282X20130151>
- McDougal, R. A., & Shepherd, G. M. (2015). 3D-printer visualization of neuron models. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00018>
- Meehan, C. F., MacDermid, V. E., Montague, S. J., Neuber-Hess, M., & Rose, P. K. (2011). Dendrite-derived supernumerary axons on adult axotomized motor neurons possess proteins that are essential for the initiation and propagation of action potentials and synaptic vesicle release. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(18), 6732–6740. <https://doi.org/10.1523/JNEUROSCI.5377-10.2011>
- Megjhani, M., Rey-Villamizar, N., Merouane, A., Lu, Y., Mukherjee, A., Trett, K., Chong, P., Harris, C., Shain, W., & Roysam, B. (2015a). Population-scale three-dimensional reconstruction and quantitative profiling of microglia arbors. *Bioinformatics (Oxford, England)*, 31(13), 2190–2198. <https://doi.org/10.1093/bioinformatics/btv109>
- Megjhani, M., Rey-Villamizar, N., Merouane, A., Lu, Y., Mukherjee, A., Trett, K., Chong, P., Harris, C., Shain, W., & Roysam, B. (2015b). Population-scale three-dimensional reconstruction and quantitative profiling of microglia arbors. *Bioinformatics*, 31(13), 2190–2198. <https://doi.org/10.1093/bioinformatics/btv109>
- Miyamae, T., Chen, K., Lewis, D. A., & Gonzalez-Burgos, G. (2017). Distinct Physiological Maturation of Parvalbumin-Positive Neuron Subtypes in Mouse Prefrontal Cortex. *The Journal of Neuroscience: The Official Journal of the*

- Society for Neuroscience*, 37(19), 4883–4902.
<https://doi.org/10.1523/JNEUROSCI.3325-16.2017>
- Mohamed, E., Paisley, C. E., Meyer, L. C., Bigbee, J. W., & Sato-Bigbee, C. (2020). Endogenous opioid peptides and brain development: Endomorphin-1 and Nociceptin play a sex-specific role in the control of oligodendrocyte maturation and brain myelination. *Glia*, 68(7), 1513–1530. <https://doi.org/10.1002/glia.23799>
- Mott, D. D., Turner, D. A., Okazaki, M. M., & Lewis, D. V. (1997). Interneurons of the Dentate–Hilus Border of the Rat Dentate Gyrus: Morphological and Electrophysiological Heterogeneity. *The Journal of Neuroscience*, 17(11), 3990–4005. <https://doi.org/10.1523/JNEUROSCI.17-11-03990.1997>
- Mottini, A., Descombes, X., & Besse, F. (2015). From Curves to Trees: A Tree-like Shapes Distance Using the Elastic Shape Analysis Framework. *Neuroinformatics*, 13(2), 175–191. <https://doi.org/10.1007/s12021-014-9255-0>
- Müller, H.-M., Rangarajan, A., Teal, T. K., & Sternberg, P. W. (2008). Textpresso for Neuroscience: Searching the Full Text of Thousands of Neuroscience Research Papers. *Neuroinformatics*, 6(3), 195–204. <https://doi.org/10.1007/s12021-008-9031-0>
- Müller, H.-M., Van Auken, K. M., Li, Y., & Sternberg, P. W. (2018). Textpresso Central: A customizable platform for searching, text mining, viewing, and curating biomedical literature. *BMC Bioinformatics*, 19(1), 94.
<https://doi.org/10.1186/s12859-018-2103-8>

- Myatt, D. R., Hadlington, T., Ascoli, G. A., & Nasuto, S. J. (2012). Neuromantic – from Semi-Manual to Semi-Automatic Reconstruction of Neuron Morphology. *Frontiers in Neuroinformatics*, *6*. <https://doi.org/10.3389/fninf.2012.00004>
- Nanda, S., Allaham, M. M., Bergamino, M., Polavaram, S., Armañanzas, R., Ascoli, G. A., & Parekh, R. (2015). Doubling up on the Fly: NeuroMorpho.Org Meets Big Data. *Neuroinformatics*, *13*(1), 127–129. <https://doi.org/10.1007/s12021-014-9257-y>
- Nanda, S., Chen, H., Das, R., Bhattacharjee, S., Cuntz, H., Torben-Nielsen, B., Peng, H., Cox, D. N., De Schutter, E., & Ascoli, G. A. (2018a). Design and implementation of multi-signal and time-varying neural reconstructions. *Scientific Data*, *5*, 170207. <https://doi.org/10.1038/sdata.2017.207>
- Nanda, S., Chen, H., Das, R., Bhattacharjee, S., Cuntz, H., Torben-Nielsen, B., Peng, H., Cox, D. N., De Schutter, E., & Ascoli, G. A. (2018b). Design and implementation of multi-signal and time-varying neural reconstructions. *Scientific Data*, *5*(1), 170207. <https://doi.org/10.1038/sdata.2017.207>
- Nave, K.-A. (2010). Myelination and the trophic support of long axons. *Nature Reviews Neuroscience*, *11*(4), 275–283. <https://doi.org/10.1038/nrn2797>
- Oga, T., Okamoto, T., & Fujita, I. (2016). Basal Dendrites of Layer-III Pyramidal Neurons do not Scale with Changes in Cortical Magnification Factor in Macaque Primary Visual Cortex. *Frontiers in Neural Circuits*, *10*, 74. <https://doi.org/10.3389/fncir.2016.00074>

- O'Halloran, D. M. (2020). Module for SWC neuron morphology file validation and correction enabled for high throughput batch processing. *PloS One*, *15*(1), e0228091. <https://doi.org/10.1371/journal.pone.0228091>
- Ohgomori, T., Yamada, J., Takeuchi, H., Kadomatsu, K., & Jinno, S. (2016a). Comparative morphometric analysis of microglia in the spinal cord of SOD1^{G93A} transgenic mouse model of amyotrophic lateral sclerosis. *European Journal of Neuroscience*, *43*(10), 1340–1351. <https://doi.org/10.1111/ejn.13227>
- Ohgomori, T., Yamada, J., Takeuchi, H., Kadomatsu, K., & Jinno, S. (2016b). Comparative morphometric analysis of microglia in the spinal cord of SOD1^{G93A} transgenic mouse model of amyotrophic lateral sclerosis. *European Journal of Neuroscience*, *43*(10), 1340–1351. <https://doi.org/10.1111/ejn.13227>
- O'Reilly, C., Iavarone, E., & Hill, S. L. (2017). A Framework for Collaborative Curation of Neuroscientific Literature. *Frontiers in Neuroinformatics*, *11*. <https://doi.org/10.3389/fninf.2017.00027>
- Osumi-Sutherland, D., Reeve, S., Mungall, C. J., Neuhaus, F., Ruttenberg, A., Jefferis, G. S. X. E., & Armstrong, J. D. (2012). A strategy for building neuroanatomy ontologies. *Bioinformatics*, *28*(9), 1262–1269. <https://doi.org/10.1093/bioinformatics/bts113>
- Parekh, R., Armañanzas, R., & Ascoli, G. A. (2015a). The importance of metadata to assess information content in digital reconstructions of neuronal morphology. *Cell and Tissue Research*, *360*(1), 121–127. <https://doi.org/10.1007/s00441-014-2103-6>

- Parekh, R., Armañanzas, R., & Ascoli, G. A. (2015b). The importance of metadata to assess information content in digital reconstructions of neuronal morphology. *Cell and Tissue Research*, 360(1), 121–127. <https://doi.org/10.1007/s00441-014-2103-6>
- Parekh, R., Armañanzas, R., & Ascoli, G. A. (2015c). The importance of metadata to assess information content in digital reconstructions of neuronal morphology. *Cell and Tissue Research*, 360(1), 121–127. <https://doi.org/10.1007/s00441-014-2103-6>
- Parekh, R., & Ascoli, G. A. (2013). Neuronal Morphology Goes Digital: A Research Hub for Cellular and System Neuroscience. *Neuron*, 77(6), 1017–1038. <https://doi.org/10.1016/j.neuron.2013.03.008>
- Parekh, R., & Ascoli, G. A. (2015a). Quantitative Investigations of Axonal and Dendritic Arbors: Development, Structure, Function, and Pathology. *The Neuroscientist*, 21(3), 241–254. <https://doi.org/10.1177/1073858414540216>
- Parekh, R., & Ascoli, G. A. (2015b). Quantitative investigations of axonal and dendritic arbors: Development, structure, function, and pathology. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 21(3), 241–254. <https://doi.org/10.1177/1073858414540216>
- Parekh, R., & Ascoli, G. A. (2015c). Quantitative investigations of axonal and dendritic arbors: Development, structure, function, and pathology. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, 21(3), 241–254. <https://doi.org/10.1177/1073858414540216>

- Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., & Huang, Z. J. (2017).
Transcriptional Architecture of Synaptic Communication Delineates GABAergic
Neuron Identity. *Cell*, *171*(3), 522-539.e20.
<https://doi.org/10.1016/j.cell.2017.08.032>
- Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., & Ascoli, G.
A. (2015). BigNeuron: Large-Scale 3D Neuron Reconstruction from Optical
Microscopy Images. *Neuron*, *87*(2), 252–256.
<https://doi.org/10.1016/j.neuron.2015.06.036>
- Peng, H., Long, F., Zhao, T., & Myers, E. (2011). Proof-editing is the Bottleneck Of 3D
Neuron Reconstruction: The Problem and Solutions. *Neuroinformatics*, *9*(2–3),
103–105. <https://doi.org/10.1007/s12021-010-9090-x>
- Peng, H., Zhou, Z., Meijering, E., Zhao, T., Ascoli, G. A., & Hawrylycz, M. (2017a).
Automatic tracing of ultra-volumes of neuronal images. *Nature Methods*, *14*(4),
332–333. <https://doi.org/10.1038/nmeth.4233>
- Peng, H., Zhou, Z., Meijering, E., Zhao, T., Ascoli, G. A., & Hawrylycz, M. (2017b).
Automatic tracing of ultra-volumes of neuronal images. *Nature Methods*, *14*(4),
332–333. <https://doi.org/10.1038/nmeth.4233>
- Peng, H., Zhou, Z., Meijering, E., Zhao, T., Ascoli, G. A., & Hawrylycz, M. (2017c).
Automatic tracing of ultra-volumes of neuronal images. *Nature Methods*, *14*(4),
332–333. <https://doi.org/10.1038/nmeth.4233>
- Petilla Interneuron Nomenclature Group, Ascoli, G. A., Alonso-Nanclares, L., Anderson,
S. A., Barrionuevo, G., Benavides-Piccione, R., Burkhalter, A., Buzsáki, G.,

- Cauli, B., Defelipe, J., Fairén, A., Feldmeyer, D., Fishell, G., Fregnac, Y., Freund, T. F., Gardner, D., Gardner, E. P., Goldberg, J. H., Helmstaedter, M., ... Yuste, R. (2008). Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews. Neuroscience*, *9*(7), 557–568.
<https://doi.org/10.1038/nrn2402>
- Plaza-Zabala, A., Sierra-Torre, V., & Sierra, A. (2017). Autophagy and Microglia: Novel Partners in Neurodegeneration and Aging. *International Journal of Molecular Sciences*, *18*(3), 598. <https://doi.org/10.3390/ijms18030598>
- Pogodalla, N., Winkler, B., & Klämbt, C. (2022). Glial Tiling in the Insect Nervous System. *Frontiers in Cellular Neuroscience*, *16*, 825695.
<https://doi.org/10.3389/fncel.2022.825695>
- Polavaram, S., & Ascoli, G. A. (2017a). An ontology-based search engine for digital reconstructions of neuronal morphology. *Brain Informatics*, *4*(2), 123–134.
<https://doi.org/10.1007/s40708-017-0062-x>
- Polavaram, S., & Ascoli, G. A. (2017b). An ontology-based search engine for digital reconstructions of neuronal morphology. *Brain Informatics*, *4*(2), 123–134.
<https://doi.org/10.1007/s40708-017-0062-x>
- Polavaram, S., Gillette, T. A., Parekh, R., & Ascoli, G. A. (2014a). Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Frontiers in Neuroanatomy*, *8*, 138. <https://doi.org/10.3389/fnana.2014.00138>

- Polavaram, S., Gillette, T. A., Parekh, R., & Ascoli, G. A. (2014b). Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Frontiers in Neuroanatomy*, 8, 138. <https://doi.org/10.3389/fnana.2014.00138>
- Polavaram, S., Gillette, T. A., Parekh, R., & Ascoli, G. A. (2014c). Statistical analysis and data mining of digital reconstructions of dendritic morphologies. *Frontiers in Neuroanatomy*, 8. <https://doi.org/10.3389/fnana.2014.00138>
- Popovic, M. A., Carnevale, N., Rozsa, B., & Zecevic, D. (2015). Electrical behaviour of dendritic spines as revealed by voltage imaging. *Nature Communications*, 6(1), 8436. <https://doi.org/10.1038/ncomms9436>
- Puckett, J. (2011). *Zotero: A guide for librarians, researchers, and educators*. Association of College and Research Libraries.
- Rasband, M. N. (2016). Glial Contributions to Neural Function and Disease. *Molecular & Cellular Proteomics: MCP*, 15(2), 355–361. <https://doi.org/10.1074/mcp.R115.053744>
- Ray, S., Chintaluri, C., Bhalla, U. S., & Wójcik, D. K. (2016). NSDF: Neuroscience Simulation Data Format. *Neuroinformatics*, 14(2), 147–167. <https://doi.org/10.1007/s12021-015-9282-5>
- Ringnér, M. (2008). What is principal component analysis? *Nature Biotechnology*, 26(3), 303–304. <https://doi.org/10.1038/nbt0308-303>
- Rocher, A. B., Crimins, J. L., Amatrudo, J. M., Kinson, M. S., Todd-Brown, M. A., Lewis, J., & Luebke, J. I. (2010). Structural and functional changes in tau mutant

- mice neurons are not linked to the presence of NFTs. *Experimental Neurology*, 223(2), 385–393. <https://doi.org/10.1016/j.expneurol.2009.07.029>
- Ropireddy, D., & Ascoli, G. A. (2011). Potential Synaptic Connectivity of Different Neurons onto Pyramidal Cells in a 3D Reconstruction of the Rat Hippocampus. *Frontiers in Neuroinformatics*, 5, 5. <https://doi.org/10.3389/fninf.2011.00005>
- Rübel, O., Prabhat, Mr., Denes, P., Conant, D., Chang, E., & Bouchard, K. (2015). *BRAINformat: A Data Standardization Framework for Neuroscience Data* [Preprint]. Neuroscience. <https://doi.org/10.1101/024521>
- Sajja, V. S. S. S., Hlavac, N., & VandeVord, P. J. (2016). Role of Glia in Memory Deficits Following Traumatic Brain Injury: Biomarkers of Glia Dysfunction. *Frontiers in Integrative Neuroscience*, 10. <https://doi.org/10.3389/fnint.2016.00007>
- Sanes, J. R., & Masland, R. H. (2015). The Types of Retinal Ganglion Cells: Current Status and Implications for Neuronal Classification. *Annual Review of Neuroscience*, 38(1), 221–246. <https://doi.org/10.1146/annurev-neuro-071714-034120>
- Sarica, A., Cerasa, A., & Quattrone, A. (2017a). Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer’s Disease: A Systematic Review. *Frontiers in Aging Neuroscience*, 9, 329. <https://doi.org/10.3389/fnagi.2017.00329>
- Sarica, A., Cerasa, A., & Quattrone, A. (2017b). Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer’s Disease: A Systematic

- Review. *Frontiers in Aging Neuroscience*, 9, 329.
<https://doi.org/10.3389/fnagi.2017.00329>
- Scorcioni, R., Polavaram, S., & Ascoli, G. A. (2008a). L-Measure: A web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, 3(5), 866–876.
<https://doi.org/10.1038/nprot.2008.51>
- Scorcioni, R., Polavaram, S., & Ascoli, G. A. (2008b). L-Measure: A web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, 3(5), 866–876.
<https://doi.org/10.1038/nprot.2008.51>
- Scorcioni, R., Polavaram, S., & Ascoli, G. A. (2008c). L-Measure: A web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols*, 3(5), 866–876.
<https://doi.org/10.1038/nprot.2008.51>
- Sebastián-Serrano, Á., Engel, T., de Diego-García, L., Olivos-Oré, L. A., Arribas-Blázquez, M., Martínez-Frailes, C., Pérez-Díaz, C., Millán, J. L., Artalejo, A. R., Miras-Portugal, M. T., Henshall, D. C., & Díaz-Hernández, M. (2016). Neurodevelopmental alterations and seizures developed by mouse model of infantile hypophosphatasia are associated with purinergic signalling deregulation. *Human Molecular Genetics*, 25(19), 4143–4156.
<https://doi.org/10.1093/hmg/ddw248>

- Sejnowski, T. J., Koch, C., & Churchland, P. S. (1988). Computational Neuroscience. *Science*, 241(4871), 1299–1306. <https://doi.org/10.1126/science.3045969>
- Senft, S. L. (2011a). A Brief History of Neuronal Reconstruction. *Neuroinformatics*, 9(2–3), 119–128. <https://doi.org/10.1007/s12021-011-9107-0>
- Senft, S. L. (2011b). A brief history of neuronal reconstruction. *Neuroinformatics*, 9(2–3), 119–128. <https://doi.org/10.1007/s12021-011-9107-0>
- Seung, H. S., & Sümbül, U. (2014). Neuronal cell types and connectivity: Lessons from the retina. *Neuron*, 83(6), 1262–1272. <https://doi.org/10.1016/j.neuron.2014.08.054>
- Shardlow, M., Ju, M., Li, M., O'Reilly, C., Iavarone, E., McNaught, J., & Ananiadou, S. (2019). A Text Mining Pipeline Using Active and Deep Learning Aimed at Curating Information in Computational Neuroscience. *Neuroinformatics*, 17(3), 391–406. <https://doi.org/10.1007/s12021-018-9404-y>
- Shekhar, K., Lapan, S. W., Whitney, I. E., Tran, N. M., Macosko, E. Z., Kowalczyk, M., Adiconis, X., Levin, J. Z., Nemesh, J., Goldman, M., McCarroll, S. A., Cepko, C. L., Regev, A., & Sanes, J. R. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. *Cell*, 166(5), 1308–1323.e30. <https://doi.org/10.1016/j.cell.2016.07.054>
- Shepherd, G. M., Marenco, L., Hines, M. L., Migliore, M., McDougal, R. A., Carnevale, N. T., Newton, A. J. H., Surles-Zeigler, M., & Ascoli, G. A. (2019). Neuron Names: A Gene- and Property-Based Name Format, With Special Reference to

- Cortical Neurons. *Frontiers in Neuroanatomy*, 13, 25.
<https://doi.org/10.3389/fnana.2019.00025>
- Silberberg, G., & Markram, H. (2007). Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*, 53(5), 735–746.
<https://doi.org/10.1016/j.neuron.2007.02.012>
- Smith, K. C., Ehlinger, D. G., & Smith, R. F. (2015). Adolescent nicotine alters dendritic morphology in the bed nucleus of the stria terminalis. *Neuroscience Letters*, 590, 111–115. <https://doi.org/10.1016/j.neulet.2015.01.056>
- Somogyi, P., Freund, T. F., & Cowey, A. (1982). The axo-axonic interneuron in the cerebral cortex of the rat, cat and monkey. *Neuroscience*, 7(11), 2577–2607.
[https://doi.org/10.1016/0306-4522\(82\)90086-0](https://doi.org/10.1016/0306-4522(82)90086-0)
- Song, F.-E., Huang, J.-L., Lin, S.-H., Wang, S., Ma, G.-F., & Tong, X.-P. (2017). Roles of NG2-glia in ischemic stroke. *CNS Neuroscience & Therapeutics*, 23(7), 547–553. <https://doi.org/10.1111/cns.12690>
- Song, I., & Dityatev, A. (2018). Crosstalk between glia, extracellular matrix and neurons. *Brain Research Bulletin*, 136, 101–108.
<https://doi.org/10.1016/j.brainresbull.2017.03.003>
- Sümbül, U., Song, S., McCulloch, K., Becker, M., Lin, B., Sanes, J. R., Masland, R. H., & Seung, H. S. (2014). A genetic and computational approach to structurally classify neuronal types. *Nature Communications*, 5(1), 3512.
<https://doi.org/10.1038/ncomms4512>

- Szabadics, J., Varga, C., Molnár, G., Oláh, S., Barzó, P., & Tamás, G. (2006). Excitatory Effect of GABAergic Axo-Axonic Cells in Cortical Microcircuits. *Science*, *311*(5758), 233–235. <https://doi.org/10.1126/science.1121325>
- Szabo, G. G., Du, X., Oijala, M., Varga, C., Parent, J. M., & Soltesz, I. (2017). Extended Interneuronal Network of the Dentate Gyrus. *Cell Reports*, *20*(6), 1262–1268. <https://doi.org/10.1016/j.celrep.2017.07.042>
- Szentágothai, J., & Arbib, M. A. (1974). Conceptual models of neural organization. *Neurosciences Research Program Bulletin*, *12*(3), 305–510.
- Takemura, S., Aso, Y., Hige, T., Wong, A., Lu, Z., Xu, C. S., Rivlin, P. K., Hess, H., Zhao, T., Parag, T., Berg, S., Huang, G., Katz, W., Olbris, D. J., Plaza, S., Umayam, L., Aniceto, R., Chang, L.-A., Lauchie, S., ... Scheffer, L. K. (2017). A connectome of a learning and memory center in the adult *Drosophila* brain. *ELife*, *6*, e26975. <https://doi.org/10.7554/eLife.26975>
- Taniguchi, H., Lu, J., & Huang, Z. J. (2013). The Spatial and Temporal Origin of Chandelier Cells in Mouse Neocortex. *Science*, *339*(6115), 70–74. <https://doi.org/10.1126/science.1227622>
- Tasic, B., Yao, Z., Graybiel, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., ... Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72–78. <https://doi.org/10.1038/s41586-018-0654-5>

- Teeters, J. L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., Denisov, G., Siegle, J. H., Olsen, S. R., Martin, C., Chun, M., Tripathy, S., Blanche, T. J., Harris, K., Buzsáki, G., ... Sommer, F. T. (2015). Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. *Neuron*, 88(4), 629–634.
<https://doi.org/10.1016/j.neuron.2015.10.025>
- The Petilla Interneuron Nomenclature Group (PING). (2008). Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature Reviews Neuroscience*, 9(7), 557–568. <https://doi.org/10.1038/nrn2402>
- Torben-Nielsen, B. (2014a). An Efficient and Extendable Python Library to Analyze Neuronal Morphologies. *Neuroinformatics*, 12(4), 619–622.
<https://doi.org/10.1007/s12021-014-9232-7>
- Torben-Nielsen, B. (2014b). An efficient and extendable python library to analyze neuronal morphologies. *Neuroinformatics*, 12(4), 619–622.
<https://doi.org/10.1007/s12021-014-9232-7>
- Ullian, E. M., Sapperstein, S. K., Christopherson, K. S., & Barres, B. A. (2001). Control of Synapse Number by Glia. *Science*, 291(5504), 657–661.
<https://doi.org/10.1126/science.291.5504.657>
- Uylings, H. B. M., & van Pelt, J. (2002). Measures for quantifying dendritic arborizations. *Network (Bristol, England)*, 13(3), 397–414.
- Van Rossum, G. (2020). *The Python Library Reference* (3.10.2) [Python].
<https://docs.python.org/3/library/random.html>

- Van Rossum, G., & Fred L., D. (2009). *Python 3 Reference Manual* (3.10.0) [Python].
<https://www.python.org/downloads/>
- Varga, C., Oijala, M., Lish, J., Szabo, G. G., Bezaire, M., Marchionni, I., Golshani, P., & Soltesz, I. (2014). Functional fission of parvalbumin interneuron classes during fast network events. *ELife*, *3*, e04006. <https://doi.org/10.7554/eLife.04006>
- Vaz, A. I. F., & Vicente, L. N. (2009). PSwarm: A hybrid solver for linearly constrained global derivative-free optimization. *Optimization Methods and Software*, *24*(4–5), 669–685. <https://doi.org/10.1080/10556780902909948>
- Veldman, M. B., Park, C. S., Eyermann, C. M., Zhang, J. Y., Zuniga-Sanchez, E., Hirano, A. A., Daigle, T. L., Foster, N. N., Zhu, M., Langfelder, P., Lopez, I. A., Brecha, N. C., Zipursky, S. L., Zeng, H., Dong, H.-W., & Yang, X. W. (2020a). Brainwide Genetic Sparse Cell Labeling to Illuminate the Morphology of Neurons and Glia with Cre-Dependent MORF Mice. *Neuron*, *108*(1), 111-127.e6.
<https://doi.org/10.1016/j.neuron.2020.07.019>
- Veldman, M. B., Park, C. S., Eyermann, C. M., Zhang, J. Y., Zuniga-Sanchez, E., Hirano, A. A., Daigle, T. L., Foster, N. N., Zhu, M., Langfelder, P., Lopez, I. A., Brecha, N. C., Zipursky, S. L., Zeng, H., Dong, H.-W., & Yang, X. W. (2020b). Brainwide Genetic Sparse Cell Labeling to Illuminate the Morphology of Neurons and Glia with Cre-Dependent MORF Mice. *Neuron*, *108*(1), 111-127.e6.
<https://doi.org/10.1016/j.neuron.2020.07.019>

- Verkhatsky, A., Ho, M. S., & Parpura, V. (2019). Evolution of Neuroglia. *Advances in Experimental Medicine and Biology*, 1175, 15–44. https://doi.org/10.1007/978-981-13-9913-8_2
- Vermaas, M., Piastra, M. C., Oostendorp, T. F., Ramsey, N. F., & Tiesinga, P. H. E. (2020). FEMfuns: A Volume Conduction Modeling Pipeline that Includes Resistive, Capacitive or Dispersive Tissue and Electrodes. *Neuroinformatics*, 18(4), 569–580. <https://doi.org/10.1007/s12021-020-09458-8>
- Vezzani, A., Pascente, R., & Ravizza, T. (2017). Biomarkers of Epileptogenesis: The Focus on Glia and Cognitive Dysfunctions. *Neurochemical Research*, 42(7), 2089–2098. <https://doi.org/10.1007/s11064-017-2271-3>
- Viney, T. J., Lasztocki, B., Katona, L., Crump, M. G., Tukker, J. J., Klausberger, T., & Somogyi, P. (2013). Network state-dependent inhibition of identified hippocampal CA3 axo-axonic cells in vivo. *Nature Neuroscience*, 16(12), 1802–1811. <https://doi.org/10.1038/nn.3550>
- von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting: Quantifying neurons and glia in human brain. *Journal of Comparative Neurology*, 524(18), 3865–3895. <https://doi.org/10.1002/cne.24040>
- Wan, Y., Long, F., Qu, L., Xiao, H., Hawrylycz, M., Myers, E. W., & Peng, H. (2015a). BlastNeuron for Automated Comparison, Retrieval and Clustering of 3D Neuron Morphologies. *Neuroinformatics*, 13(4), 487–499. <https://doi.org/10.1007/s12021-015-9272-7>

- Wan, Y., Long, F., Qu, L., Xiao, H., Hawrylycz, M., Myers, E. W., & Peng, H. (2015b). BlastNeuron for Automated Comparison, Retrieval and Clustering of 3D Neuron Morphologies. *Neuroinformatics*, *13*(4), 487–499. <https://doi.org/10.1007/s12021-015-9272-7>
- Wan, Y., Long, F., Qu, L., Xiao, H., Hawrylycz, M., Myers, E. W., & Peng, H. (2015c). BlastNeuron for Automated Comparison, Retrieval and Clustering of 3D Neuron Morphologies. *Neuroinformatics*, *13*(4), 487–499. <https://doi.org/10.1007/s12021-015-9272-7>
- Wang, X., Tucciarone, J., Jiang, S., Yin, F., Wang, B.-S., Wang, D., Jia, Y., Jia, X., Li, Y., Yang, T., Xu, Z., Akram, M. A., Wang, Y., Zeng, S., Ascoli, G. A., Mitra, P., Gong, H., Luo, Q., & Huang, Z. J. (2019). Genetic Single Neuron Anatomy Reveals Fine Granularity of Cortical Axo-Axonic Cells. *Cell Reports*, *26*(11), 3145-3159.e5. <https://doi.org/10.1016/j.celrep.2019.02.040>
- Wheeler, D. W., White, C. M., Rees, C. L., Komendantov, A. O., Hamilton, D. J., & Ascoli, G. A. (2015). Hippocampome.org: A knowledge base of neuron types in the rodent hippocampus. *ELife*, *4*, e09960. <https://doi.org/10.7554/eLife.09960>
- Wittner, L., & Miles, R. (2007). Factors defining a pacemaker region for synchrony in the hippocampus. *The Journal of Physiology*, *584*(Pt 3), 867–883. <https://doi.org/10.1113/jphysiol.2007.138131>
- Woodruff, A. R., Anderson, S. A., & Yuste, R. (2010). The Enigmatic Function of Chandelier Cells. *Frontiers in Neuroscience*, *4*. <https://doi.org/10.3389/fnins.2010.00201>

- Xiong, H., Zhou, Z., Zhu, M., Lv, X., Li, A., Li, S., Li, L., Yang, T., Wang, S., Yang, Z., Xu, T., Luo, Q., Gong, H., & Zeng, S. (2014). Chemical reactivation of quenched fluorescent protein molecules enables resin-embedded fluorescence microimaging. *Nature Communications*, 5(1), 3992. <https://doi.org/10.1038/ncomms4992>
- Yamawaki, N., Borges, K., Suter, B. A., Harris, K. D., & Shepherd, G. M. G. (2014). A genuine layer 4 in motor cortex with prototypical synaptic circuit connectivity. *ELife*, 3, e05422. <https://doi.org/10.7554/eLife.05422>
- Yang, J., He, Y., & Liu, X. (2020). Retrieving similar substructures on 3D neuron reconstructions. *Brain Informatics*, 7(1), 14. <https://doi.org/10.1186/s40708-020-00117-x>
- Yang, J., Yang, H., Liu, Y., Li, X., Qin, L., Lou, H., Duan, S., & Wang, H. (2016). Astrocytes contribute to synapse elimination via type 2 inositol 1,4,5-trisphosphate receptor-dependent release of ATP. *ELife*, 5, e15043. <https://doi.org/10.7554/eLife.15043>
- Yang, T., Zheng, T., Shang, Z., Wang, X., Lv, X., Yuan, J., & Zeng, S. (2015). Rapid imaging of large tissues using high-resolution stage-scanning microscopy. *Biomedical Optics Express*, 6(5), 1867. <https://doi.org/10.1364/BOE.6.001867>
- Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., & Linnarsson, S. (2015). Cell types in the mouse

- cortex and hippocampus revealed by single-cell RNA-seq. *Science*, *347*(6226), 1138–1142. <https://doi.org/10.1126/science.aaa1934>
- Zeng, H., & Sanes, J. R. (2017). Neuronal cell-type classification: Challenges, opportunities and the path forward. *Nature Reviews Neuroscience*, *18*(9), 530–546. <https://doi.org/10.1038/nrn.2017.85>
- Zhang, T., Zeng, Y., Zhang, Y., Zhang, X., Shi, M., Tang, L., Zhang, D., & Xu, B. (2021a). Neuron type classification in rat brain based on integrative convolutional and tree-based recurrent neural networks. *Scientific Reports*, *11*(1), 7291. <https://doi.org/10.1038/s41598-021-86780-4>
- Zhang, T., Zeng, Y., Zhang, Y., Zhang, X., Shi, M., Tang, L., Zhang, D., & Xu, B. (2021b). Neuron type classification in rat brain based on integrative convolutional and tree-based recurrent neural networks. *Scientific Reports*, *11*(1), 7291. <https://doi.org/10.1038/s41598-021-86780-4>
- Zhou, B., Zuo, Y., & Jiang, R. (2019). Astrocyte morphology: Diversity, plasticity, and role in neurological diseases. *CNS Neuroscience & Therapeutics*, *25*(6), 665–673. <https://doi.org/10.1111/cns.13123>
- Zippo, A. G., & Biella, G. E. M. (2015). Quantifying the Number of Discriminable Coincident Dendritic Input Patterns through Dendritic Tree Morphology. *Scientific Reports*, *5*(1), 11543. <https://doi.org/10.1038/srep11543>
- Zisis, E., Keller, D., Kanari, L., Arnaudon, A., Gevaert, M., Delemontex, T., Coste, B., Foni, A., Abdellah, M., Calì, C., Hess, K., Magistretti, P. J., Schürmann, F., & Markram, H. (2021a). Digital Reconstruction of the Neuro-Glia-Vascular

Architecture. *Cerebral Cortex*, 31(12), 5686–5703.

<https://doi.org/10.1093/cercor/bhab254>

Zisis, E., Keller, D., Kanari, L., Arnaudon, A., Gevaert, M., Delemontex, T., Coste, B.,

Foni, A., Abdellah, M., Calì, C., Hess, K., Magistretti, P. J., Schürmann, F., &

Markram, H. (2021b). Digital Reconstruction of the Neuro-Glia-Vascular

Architecture. *Cerebral Cortex (New York, N.Y.: 1991)*, 31(12), 5686–5703.

<https://doi.org/10.1093/cercor/bhab254>

Zuchero, J. B., & Barres, B. A. (2015). Glia in mammalian development and disease.

Development, 142(22), 3805–3809. <https://doi.org/10.1242/dev.129304>

BIOGRAPHY

Masood A. Akram graduated from DJ Sindh Government High School, Karachi, Pakistan, in 2007. He received his Doctor of Pharmacy from University of Karachi in 2012. He was employed as a research assistant in the Center for Neural Informatics, Structures, and Plasticity (CN3) of George Mason University for 5 years and 8 months and received his Doctor of Philosophy in Bioengineering from George Mason University, Fairfax, USA in 2022.